Calhoun
Institutional Archive of the Naval Postgraduate School

Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations 1. Thesis and Dissertation Collection, all items

1983

A preliminary analysis of human factors affecting the recognition accuracy of a discrete word recognizer for C3 systems.

Yellen, Howard William.

http://hdl.handle.net/10945/19948

Downloaded from NPS Archive: Calhoun

DUDLEY KNOX LIBRARY

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

IBR.RY, NAVAL POSTGRADUATE SCHOOL
ONTEREY, CA 93940

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A PRELIMINARY ANALYSIS OF HUMAN FACTORS AFFECTING THE RECOGNITION ACCURACY OF A DISCRETE WORD RECOGNIZER FOR C3 SYSTEMS

by

Howard William Yellen

March 1983

Thesis Advisor: G. K. Poock

Approved for public release; distribution unlimited.

T207864

LIBRARY, NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93940

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Preliminary Analysis of Human Factors Affecting the Recognition Accuracy of a Discrete Word Recognizer for C3 Systems		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; March 1983
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Howard William Yellen		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1983
		13. NUMBER OF PAGES 190
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Voice Recognition
Human Factors
Automatic Speech Recognition
Statistical Significance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Literature pertaining to Voice Recognition abounds with information relevant to the assessment of transitory speech recognition devices. In the past, engineering requirements have dictated the path this technology followed. But, other factors do exist that influence recognition accuracy. This thesis explores the impact of Human Factors on the successful recognition of speech, principally addressing the differences or variability among users. A Threshold Technology T-600

was used for a 100 utterance vocabulary to test 44 subjects. A statistical analysis was conducted on 5 generic categories of Human Factors: Occupational, Operational, Psychological, Physiological and Personal. How the equipment is trained and the experience level of the speaker were found to be key characteristics influencing recognition accuracy. To a lesser extent computer experience, time of week, accent, vital capacity and rate of air flow, speaker cooperativeness and anxiety were found to affect overall error rate.

Approved for public release; distribution unlimited.

A Preliminary Analysis of Human Factors Affecting The Recognition Accuracy of a Discrete Word Recognizer For C3 Systems

by

Howard William Yellen
Captain, United States Army
B.A., Temple University, 1972

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS TECHNOLOGY
(COMMAND, CONTROL, AND COMMUNICATIONS)

from the

NAVAL POSTGRADUATE SCHOOL
March 1983

Y345
c.1

ABSTRACT

Literature pertaining to Voice Recognition abounds with information relevant to the assessment of transitory speech recognition devices. In the past, engineering requirements have dictated the path this technology followed. But, other factors do exist that influence recognition accuracy. This thesis explores the impact of Human Factors on the successful recognition of speech, principally addressing the differences or variability among users. A Threshold Technology T-600 was used for a 100 utterance vocabulary to test 44 subjects. A statistical analysis was conducted on 5 generic categories of Human Factors: Occupational, Operational, Psychological, Physiological and Personal. How the equipment is trained and the experience level of the speaker were found to be key characteristics influencing recognition accuracy. To a lesser extent computer experience, time of week, accent, vital capacity and rate of air flow, speaker cooperativeness and anxiety were found to affect overall error rates.

TABLE OF CONTENTS

I. INTRODUCTION 14

II. COMPUTER RECOGNITION OF SPEECH 18

- A. OVERVIEW OF VOICE INPUT TECHNOLOGY 18
- B. THE VALUE OF SPEECH RECOGNITION 26
  - 1. Advantages of Speech Recognition 27
  - 2. Limitations of Speech Recognition 29
- C. APPLICABILITY OF COMPUTER RECOGNITION OF SPEECH 32
  - 1. Commercial Applications 32
  - 2. Military Applications 34

III. HUMAN FACTORS IN SPEECH RECOGNITION 40

- A. DEFINITION AND PURPOSE 40
- B. FACTORS AFFECTING RECOGNITION ACCURACY 41
  - 1. General 41
  - 2. Differences Between Speakers 44
  - 3. Differences Within Speakers 48
  - 4. Miscellaneous Factors 50

IV. DESCRIPTION OF THE EXPERIMENT 53

- A. OBJECTIVES AND CONSTRAINTS 53
  - 1. Objectives 53
    - a. Occupational Characteristics 53
    - b. Operational Characteristics 54
    - c. Personal Characterisitcs 55

d. Physiological Characteristics.......... 56

e. Psychological Characteristics.......... 57

2. Constraints.................................. 58

B. SUBJECTS..................................... 59

C. EQUIPMENT.................................... 60

1. Voice Recognition System..................... 60

2. Spirometer................................... 67

3. Peak Flow Meter.............................. 70

4. Tape Recorder................................ 72

D. INSTRUMENTATION.............................. 73

1. User Questionnaire #1........................ 74

2. User Questionnaire #2........................ 74

3. STAI Questionnaire........................... 75

E. EXPERIMENTAL DESIGN.......................... 76

F. PROCEDURE.................................... 76

1. Training..................................... 76

2. Recognition Testing.......................... 79

3. Vocabulary................................... 80

G. VARIABLES.................................... 80

V. ANALYSIS AND RESULTS......................... 82

A. GENERAL...................................... 82

B. OCCUPATIONAL CHARACTERISTICS................. 84

1. Hypotheses................................... 84

2. Job Function................................. 85

3. Branch of Service............................ 87

4. Job and Serice Satisfaction.................. 89

5. Previous Computer Experience............... 91
6. Foreign Language Competency................ 93
C. OPERATIONAL CHARACTERISTICS.................. 94
1. Hypotheses................................. 94
2. Method of Training......................... 95
3. Time of Day and Week....................... 97
4. User Experience............................ 98
5. Ease of Use................................101
D. PERSONAL CHARACTERISTICS.....................102
1. Hypotheses.................................102
2. Race.......................................104
3. Marital Status and Family Size.............104
4. Religious Preference.......................106
5. Accent.....................................107
6. Place of Birth and Geographic Origin.......108
7. Level of Education.........................110
8. Socio-economic Class.......................112
9. Dental.....................................113
E. PHYSIOLOGICAL CHARACTERISTICS................114
1. Hypotheses.................................114
2. Age..115
3. Height and Weight..........................116
4. Vital Capacity and Rate of Air Flow........118
5. Physical Condition.........................122

F. PSYCHOLOGICAL CHARACTERISTICS....................124

1. Hypotheses..................................124

2. Psychological Anxiety.......................124

3. Speaker Cooperativeness....................129

4. Recognition Errors...........................131

5. Attitudes Toward The Use of Voice..........132

6. Attitude Toward Computers and Information Processing..................................136

G. VOCABULARY ERRORS..............................138

VI. CONCLUSIONS...................................141

APPENDIX A: USER QUESTIONNAIRE #1...................147

APPENDIX B: USER QUESTIONNAIRE #2...................156

APPENDIX C: SELF-EVALUATION QUESTIONNAIRE...........161

APPENDIX D: SELF-EVALUATION QUESTIONNAIRE...........164

APPENDIX E: UTTERANCE LIST: TRAINING WEEK - WEEK#1.....167

APPENDIX F: UTTERANCE LIST: WEEK#2...................170

APPENDIX G: UTTERANCE LIST: WEEK#3...................173

APPENDIX H: DATA COLLECTION FORM.....................176

APPENDIX I: MASTER LIST OF UTTERANCES...............181

APPENDIX J: INDIVIDUAL SUBJECT RECOGNITION RATES........184

LIST OF REFERENCES.................................186

INITIAL DISTRIBUTION LIST..........................189

LIST OF FIGURES

1. Speech Recognition Model............................ 20
2. Processing Functions of a Speech Recognition System.. 23
3. T-600 Speech Recognition Equipment.................. 63
4. Acoustic Sound Reduction Chamber.................... 64
5. Placement of the SHURE SM-10 Microphone............. 65
6. Recording Spirometer................................ 68
7. Use of Recording Spirometer to Measure and Record Vital Capacity....................................... 69
8. The Wright Peak Flow Meter.......................... 71
9. Measurement of Speaker's Rate of Air Flow........... 71
10. AKAI Tape Recorder.................................. 72
11. Experimental Design................................. 77
12. Mean Error Rate vs. Job Function.................... 85
13. Mean Error Rate vs. Branch of Service............... 89
14. Mean Error Rate vs. Computer Experience............. 93
15. Mean Error Rate vs. Training Method................. 96
16. Trials versus Job Function..........................100
17. Trials versus Training Method.......................100
18. Mean Error Rate versus Accent.......................108
19. Mean Error Rate vs. Education.......................112
20. Mean Error Rate vs. Vital Capacity..................120
21. Mean Error Rate vs. Rate of Air Flow................120

22. Scatter Plot for Vital Capacity........................121

23. Scatter Plot for Rate of Air Flow.....................121

24. Mean Error Rate vs. State Anxiety (Week #1).........126

25. Mean Error Rate vs. State Anxiety (Week #2) 126

26. Mean Error Rate vs. Trait Anxiety 127

27. Mean Error Rate vs. Speaker Cooperativeness.........130

28. Scatter Plot: Mean Error Rate vs. Question #4......134

29. Scatter Plot: Mean Error Rate vs. Question #6 134

30. Scatter Plot: Mean Error Rate vs. Question #8 135

31. Mean Error Rate vs. # Syllables (by Week)...........139

32. Mean Error Rate vs. # Syllables (Overall)...........140

LIST OF TABLES

I. MILITARY APPLICATIONS FOR SPEECH RECOGNITION.... 35

II. DIMENSIONS OF DIFFICULTY FOR SPEECH RECOGNITION.................................... 43

III. SUBJECT CHARACTERISTICS........................... 61

IV. TEST FOR EQUALITY OF VARIANCES.................. 83

V. ANALYSIS OF VARIANCE FOR RECOGNITION ACCURACY... 86

VI. MEAN TOTAL ERROR RATES FOR JOB FUNCTION BY WEEKS.. 87

VII. AFFECT BY BRANCH OF SERVICE....................... 88

VIII. AFFECT BY JOB/SERVICE SATISFACTION.............. 90

IX. AFFECT OF COMPUTER EXPERIENCE.................... 92

X. AFFECT OF COMPETENCY IN ANOTHER LANGUAGE........ 94

XI. MEAN TOTAL ERROR RATES FOR METHOD OF TRAINING BY WEEKS.. 96

XII. AFFECT OF TIME OF DAY AND WEEK.................. 98

XIII. AFFECT DUE TO USER EXPERIENCE.................... 99

XIV. AFFECT DUE TO EASE OF USE OF VOICE EQUIPMENT....102

XV. AFFECT OF RACE ON RECOGNITION ACCURACY..........104

XVI. AFFECT OF MARITAL STATUS AND FAMILY SIZE........105

XVII. AFFECT OF RELIGIOUS PREFERENCE..................106

XVIII. AFFECT OF ACCENT ON RECOGNITION ACCURACY........107

XIX. AFFECT OF PLACE OF BIRTH AND GEOGRAPHIC ORIGIN..109

XX. AFFECT OF LEVEL OF EDUCATION......................111

XXI. AFFECT OF SOCIO-ECONOMIC CLASS...................113

XXII. AFFECT OF PAST AND/OR PRESENT DENTAL CARE.......114

XXIII. AFFECT ON RECOGNITION ACCURACY DUE TO AGE.......116

XXIV. AFFECT OF HEIGHT AND WEIGHT ON RECOGNITION ACCURACY....................................117

XXV. AFFECT OF VITAL CAPACITY AND RATE OF AIR FLOW...119

XXVI. AFFECT ON RECOGNITION ACCURACY DUE TO PHYSICAL CONDITION................................123

XXVII. AFFECT ON RECOGNITION ACCURACY DUE TO ANXIETY...128

XXVIII. AFFECT OF SPEAKER COOPERATION AND PARTICIPATION..................................130

XXIX. AFFECT OF RECOGNITION ERRORS......................132

XXX. AFFECT DUE TO ATTITUDES PERTAINING TO THE USE OF VOICE......................................133

XXXI. AFFECT DUE TO ATTITUDES TOWARD COMPUTERS AND DATA PROCESSING............................137

ACKNOWLEDGEMENTS

I wish to express my thanks to my thesis advisor, Professor Gary Poock for introducing me to the world of voice technology, allowing me the independence to conduct the experimentation as I desired, and for the competitive challenge posed on the racquetball court; to CDR Chuck Hutchins for his expertise and advice in Human Factors and for serving as second reader; to Jay Martin and Ellen Roland for their practical advice; and to Paul Sparks for his technical assistance and advice.

Finally, my sincerest thanks to my wife, Susan for her help, understanding and encouragement; and to my son, Michael, who has spent the better part of three months wondering where Dad was, for his special smile and big hug when it was needed the most.

I. INTRODUCTION

The insistence and dependence upon state of the art equipment has been a predominant characteristic throughout the efforts within the Command and Control community. Despite the penchant for newer, better, and more sophisticated equipment, there must exist some measure of emphasis on the personnel needed to train with, operate on, and maintain the readiness of, such equipment. Personnel considerations cannot be divorced from test programs designed to identify optimal systems or equipment. When these considerations are carefully examined, then the data obtained from such programs can be effectively used to enhance personnel subsystem design and implementation.

A personnel subsystem test program is one which places the requisite emphasis on personnel rather than equipment. Kryter [Ref. 1] enumerates six objectives necessary for a successful test program.

1. To evaluate whether the system can be operated, maintained and controlled by the personnel assigned to it.

2. To determine the effect of human performance on system performance and vice versa. This objective is aimed at discovering critical inadequacies in man-machine

interaction and subsequently identify changes that would improve their compatibility.

3. To develop valid qualitative and quantitative personnel requirements, selection procedures, and tables of organizational manning. How many and what type of people will provide optimal effectiveness of the man-machine interface?

4. To evaluate individual and/or long term operational readiness and applicable training programs.

5. To evaluate training equipment and supporting materials.

6. To evaluate job aids, technical publications and other tools for training and for assisting on the job performance.

Increased productivity through automation involves two major issues; technological and human. Speech is a uniquely human capability. Speech recognition by a computer involves getting a machine to accept, recognize, and correctly respond to spoken messages. This machine must take the input speech, compare it against the expected pronunciation for allowable utterances, identify the intended message or utterance, and produce the correct and appropriate response. To adequately implement the capabilities of such a technology, the objectives above become all the more

relevant. Of paramount importance is the human, for it takes people to make all this automation work.

Speech recognizers commercially available today are effective only within narrow limits. They have relatively small vocabularies and 'frequently' confuse words. Within this context, it becomes incumbent upon the user to develop the skill to talk to the recognizer [Ref. 2: p. 26]. As such, a recognizer's performance will vary widely from speaker to speaker.

Much of the work in speech recognition has centered on the development and improvement of speech recognition devices. For example:

-- Linear Predictive Coding (LPC) in early '70s

-- Dynamic programming

-- Development of 1 million bit/sec processors

A user's experience notwithstanding, the human variable in recognition performance remains strong. This has often been observed in the past and even led to a description of user categories [Ref. 2: p. 30] of 'sheeps' and 'goats'. These speech recognition systems work well for the 'sheep' but the majority of the problems are created by a small segment of the population - the 'goats'.

Recognizing the significant impact that engineers have had on perpetuating the continued advent and technological advancement of speech recognition, it is nevertheless,

critical to remind ourselves of the interdisciplinary nature of speech recognition. Besides engineering, the total discipline of speech sciences and technology includes such traditional disciplines as psychology, linguistics, anatomy and physiology, computer sciences and human factors. This thesis endeavors to examine the impact of human factors on the successful recognition of speech, principally addressing the differences or variability among users.

First, the modality of voice input will be examined citing some of the more readily apparent advantages and disadvantages, and an overview provided as to its potential applicability in a Command and Control environment. With a general appreciation of speech recognition (the term ′voice recognition′ is synonomous and used interchangeably within this document) in hand, the variety of human factors that can affect the successful recognition of speech by a machine will then be summarized. Subsequently, the experimental methodology used to examine and differentiate speech recognition equipment users will be presented. Lastly, the experimental results will be presented and an analysis provided of the correlation of each variable examined to its associated error rates as well as an analysis of variance.

II. COMPUTER RECOGNITION OF SPEECH

A. OVERVIEW OF VOICE INPUT TECHNOLOGY

Speech recognition can be considered as a subset of a broader field known as Speech Understanding. Speech Understanding Systems (SUS) have the objective of interpreting the intent of the speaker whether or not the user's speech is grammatically correct or well formed. While Speech Recognition Systems (SRS) are primarily interested in the correct recognition of every word, SUS are concerned with the meaning of entire conversational segments.

Until now the only significant undertaking has been the ARPA SUR project [Ref. 3], a five year effort with the objective of obtaining a breakthrough in speech understanding capability that would then allow the development of practical man-machine communication systems. Specifically, the objectives were to develop a SUS that would accept continuous speech from many cooperative speakers of a general American public; a system which used syntactic analysis, semantics, pragmatic information and prosodics to acquire an appropriate computer response.

The goals of speech recognition, in contrast, are less ambitious. Instead of abstract concepts such as meaning or understanding, SRS try to solve the more practical problems

of analyzing the acoustic waveform and applying pattern recognition techniques in order to differentiate between utterances [Ref. 4]. Figure 1 illustrates a typical speech recognition model.

The acoustic speech signal is first analyzed to extract such acoustic parameters as frequency spectrum and the energy in different time segments. Next, information carrying features are extracted that define various phonetic events such as how noisy (fricative-like) the signal is, positions of different vowel-like sounds and vibration of the speaker's vocal cords. This information is then used to divide the speech into time slices or segments and are labelled with phonetic categories. The phonetic sequence for the input speech is matched to stored sequences of expected pronunciations for the words in the lexicon or dictionary, and the best matching sequences are determined to be the most likely word(s) that had occurred in speech.

Speech recognition systems can be considered as belonging to one of two categories; continuous (connected) or isolated (discrete) speech systems. Continuous systems are those which can extract information from strings of words even though the words run together as in natural speech. Isolated systems require a short pause before and after utterances that are to be recognized as entities. The minimum duration of a pause is typically between 100-200 msec. An isolated word recognizer is also limited in the

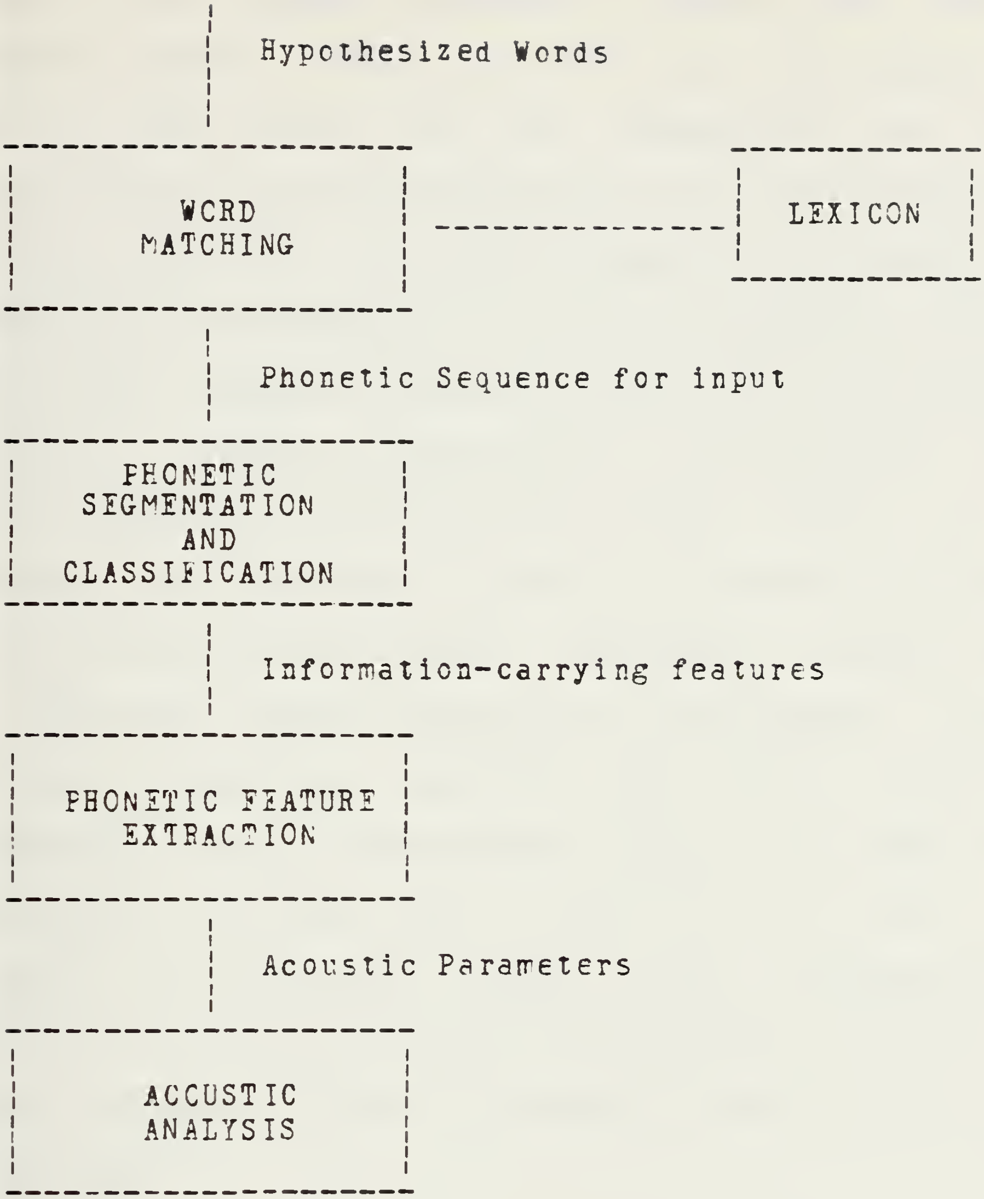
Hypothesized Words
WORD MATCHING
LEXICON
Phonetic Sequence for input
PHONETIC SEGMENTATION AND CLASSIFICATION
Information-carrying features
PHONETIC FEATURE EXTRACTION
Acoustic Parameters
ACOUSTIC ANALYSIS

Figure 1. Speech Recognition Model
(From Reference 4)

duration of the spoken utterance, usually 2-4 seconds. Continuous speech recognizers are just now beginning to appear on the market but are expensive and their capabilities and reliability have yet to be realistically or practically evaluated. For the remainder of this thesis our discussion will be confined to discrete recognition systems.

Two other concepts of speech recognition to be discussed are that of speaker independence and vocabulary size. Speaker dependent systems are those which require speaker adaptation (or 'training') in order to achieve recognition. This is in contrast to speaker independent systems which will recognize speech regardless of the speaker. In terms of speech recognition equipment and their associated vocabularies, most recognizers work well with small vocabularies of 10-50 words [Ref. 5: p. 80]. The possibility of confusion between words increases as the vocabulary size increases, and to some extent the chance of similar sounding words increases with such larger vocabularies.

At this juncture it is appropriate to expand our definition of 'words' to encompass more than just individual words. As used herein, 'word' is used interchangeably with the term 'utterance' and may be either a singular mono- or polysyllabic word or a combination of mono- or polysyllable words joined into a phrase. (ie. Place-a-Circle-on-Moscow)

The four processing functions [Ref. 6] contained in a limited vocabulary voice recognition system, as shown in Figure 2, consist of a transducer, preprocessor, feature extractor, and a final decision-level classifier.

1. Transducer: The microphone is the interface between the user and the system and converts the spoken phrase into electrical signals that are analyzed by the other components of the system.

2. Preprocessor: No matter how it is represented, spectral information must be explicitly or implicitly contained in all speech encodings. The initial analyses produce parametric representations [Ref. 7] and take place in the preprocessor. This segment of the system transforms the speech signal in order to enhance certain properties and make them more easily detectable in a speech recognition system. The signal is normalized in time by dynamic programming for subsequent comparisons with various reference patterns. Data Compression removes any extraneous or irrelevant information. Both time and frequency domain analytical techniques are performed on the input signal. Speech analysis is achieved by either direct analog spectrum analysis via fast fourier transform (FFT) in the frequency domain, or linear predictive coding (LPC) in the time domain.

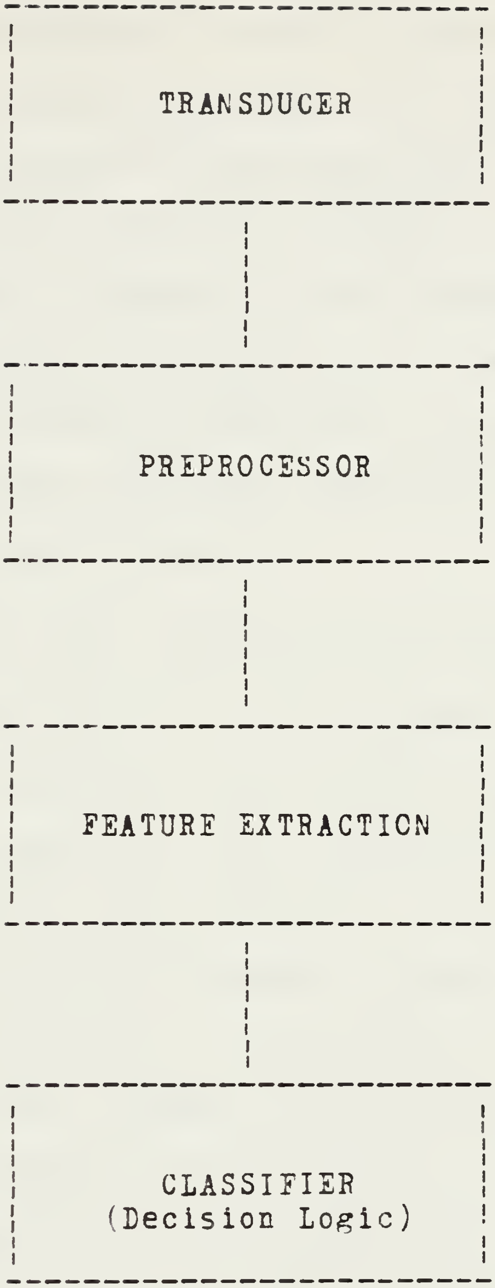

Figure 2. Processing Functions of a Speech Recognition System (From Reference 6)

3. Feature Extraction: The key processing function in a pattern recognition system is the feature extractor. The more optimal the set of acoustical features extracted and sent to the classifier, the less complex the classifier need be to achieve a given accuracy level. This segment of the system produces a set number of significant acoustical features (depending on the individual recognizer) a few of which include spectral slopes, phonetic classification, and initial estimate of word boundary.

4. Classifier: The classification process is performed in software using a minicomputer. When a speaker issues an utterance, the encoded features and their time of occurrence are stored in short term memory. The duration of the utterance is broken into time segments and the features reconstructed into the normalized time base. Reference patterns, previously input by the speaker for the system's vocabulary of words are compared to the feature occurrence patterns and a 'best-fit' or 'closest-match' determined for a word decision. The number of bits of information for the feature map of each reference pattern is determined by mapping the number of acoustic features onto the number of time segments.

The first two processing functions are accomplished by a hard wired preprocessor and feature extractor. This achieves real-time processing since only the classification function is performed in a general-purpose minicomputer [Ref. 6: p. 177].

A discrete word recognizer must be 'trained' for individual talkers and/or words. This can be done by a user simply speaking a set number of training samples into the device to provide a reference set of features. The system stores in memory the reference set of word features for each word (utterance) the user has spoken. Once the system is trained, the user may speak words into the device during normal operation and these are compared with the stored patterns. The 'closest fit' is selected as the recognized word. This sequence of events is commonly partitioned into the training and recognition modes of operation.

There are two types of errors that can occur in speech recognition. The first is a rejection, or the inability of the recognizer to correctly classify an utterance. The second, and in a practical sense more troublesome, is a misrecognition. This occurs when the recognizer classifies an utterance as something other than what was spoken. Better recognizers usually have recognition algorithms designed to reject rather than guess at questionable words. Higher quality systems such as Threshold (Models 600 and 680) have error rates that are quite acceptable [Ref. 8, 9,

10]. Extensive experimentation has shown approximate error rates to be between .2 and 11.4 percent [Ref. 6: pp. 179-180]. Of course, what constitutes an acceptable error rate is critically dependent upon the particular application and data entry rate.

B. THE VALUE OF SPEECH RECOGNITION

The Department of Defense has been very active in the past few years in their efforts to assess the merits of voice recognition with machines. Such locations as the Naval Postgraduate School, Wright Patterson Air Force Base, Rome Air Development Center, Naval Air Development Center and assorted other agencies and contractors, have conducted extensive tests in order to examine human interaction with machines through the use of voice input and other modalities. In order to comprehend the need for further research pertaining to voice input technology, it is essential to review the advantages and limitations that this type of technology offers. More importantly, it is essential to understand its potential capabilities and applications in a military environment. Is speech recognition beneficial (considering costs of $300 - $80,000+), practical, and usable to justify the continued expenditures of research and development funds (6.1 and 6.4) and operational monies.

1. Advantages of Speech Recognition

Proponents of computer recognition of speech will continually extol the virtues and unlimited possibilities the technology offers. In an abbreviated fashion, the five general advantages of voice input to machines may be summarized as follows:

- -- Natural communication
- -- Training
- -- Multimodal communication
- -- Fast communication
- -- Error reduction in data input

Speech is our most natural mode of communication. It is a familiar, spontaneous and convenient method of expressing one's thoughts, ideas, or intentions. Untrained users of voice recognition systems, regardless of whether they can read, write, type or keypunch, can all speak or make sounds. These characteristics of the speech input modality make it applicable for users at all general skill levels, from systems engineers to computer operators to blue collar workers on an assembly line.

A user of speech recognition equipment requires little or no training. They have only to restrict their spoken utterances to those which the machine can recognize. In the case of discrete systems, isolated words are separated by a short pause so as to ease the location of

word boundaries and word choices to which the machine has been trained to recognize. Although this appears to be disadvantageous, it is more realistically a compromise to natural speech in that no adverse affects are caused the user in terms of operating the speech recognition equipment.

Experimentation [Ref. 11: p. 608] has shown that speech, instead of interrupting communications necessary to perform other tasks, can enable users to do these tasks simultaneously with voice and thereby reduce or at a minimum, not add to the time required to perform a complex task. The advantage of having one's hands and eyes free to do other tasks is perhaps the pivotal point in the determination of applicability of speech recognition devices. This multimodal aspect allows us to place the microphone anywhere (headset mounted, hand-held, on a stand) and still communicate commands and information. Threshold Technology even has a wireless microphone [Ref. 12] that permits extensive mobility while talking to computers.

The fastest modality for communications by a human is speech. An individual can speak twice as fast as the average typist can type [Ref. 5: p. 45]. This has been clearly demonstrated by Ochman and Chapanis [Ref. 11] whose experimental results showed that communication via typewriter or handwriting could not approach speech in terms of speed or task efficiency. Further substantiation from the Naval Postgraduate School [Ref. 8: p. 2] showed that

voice entry was 17% faster than typing, after only three hours of training. Additionally, while speech recognition accuracy is slightly degraded by mental or motor loading of the user [Ref. 13: p. 32], voice is nevertheless faster and more accurate than other input modes when the user must perform another task while simultaneously interacting with the speech recognition equipment [Ref. 8: p. 2]

By now it is clear that speech recognition permits data entry directly into the computer without intermediate steps such as manual transcription or keypunching which are subject to error. Again, research at the Naval Postgraduate School has shown that 183% more errors occurred in manual data manipulation (typing) than by voice [Ref. 8 p. 2]. Such common entry errors as the transposition of digits, which are usually caused by eye movement or other distractions, are almost eliminated with the use of automatic speech recognition [Ref. 14].

2. Limitations of Speech Recognition

If a particular technology was devoid of errors or practical limitations, we could assure universal application and implementation. Although the advantages of speech recognition are seemingly well established, there do exist several problems associated with the ability to speak to machines. These limitations include:

-- User variability

-- Constrained speech

-- Isolated speech

-- Breath noise

-- User confusion

-- Environmental factors

Speakers exhibit a wide range of personal characteristics that add a significant measure of difficulty in the ability of a machine to recognize speech. A speaker's sex, geographic origin, and articulation experience are just a few of the elements that result in a user's variability. Consistency is also a key element in successful recognition accuracy. A speaker may talk quite differently in training the machine as compared to when he or she may use it in a practical application. Additionally, physical changes in the speaker such as age, physical condition, stress (physical or emotional), or fatigue, to name a few, can induce variability that will ultimately affect successful recognition accuracy.

An isolated word recognition system imposes a restricted (constrained) vocabulary both in terms of size and content, upon the user. This becomes a limitation when we consider that most people are accustomed to speaking in natural, fluent prose. Because of the limited vocabulary, users must be careful of the types of words included for recognition. The similarity of sound structures between words (ie. Nine vs. Time) adds a measure of confusion that can subsequently affect overall performance. Design of

a vocabulary for a particular application is an important and controllable factor in determining the acceptability of voice input for a given task.

Because isolated word recognizers depend significantly upon the detection of a minimum pause between words, word boundary detection becomes perhaps the single most critical limitation. The usual method is to measure changes in energy levels [Ref. 5]. An isolated word is detected at a point where the energy in the acoustic signal rises above a certain threshold. At the end of the word, the energy drops, and the resultant silence indicates that the utterance is over. But, energy fluctuations are not enough to detect all word boundaries, and thus advanced detection techniques will have to involve detection and inclusion of stop consonants within words, while eliminating pauses due to ′lip-smacks′ or breath noise.

In a limited vocabulary, isolated word recognition system, breath noise can be a serious problem [Ref. 6: p. 174]. An individual who is involved in little or no physical movement while engaged with a voice recognition system can achieve very high recognition accuracy. This accuracy can soon deteriorate once the user begins to move around. Inhaling will not cause any adverse affects when using a close-talking, noise-cancelling microphone, but exhaling will produce signal levels comparable to speech levels. As physical activity increases so does one′s

breathing pattern and as a result increased exhalation will lead to the above mentioned deterioration in recognition accuracy.

While voice input provides multimodal communications, this particular advantage has an inherent limitation in that the user can become confused as to what mode to use. As a result, input modalities can become confused, and interfere with each other so that the total rate of information transfer may not be as high as the sum of the rates possible with each separate modality.

Finally, the environment in which the speech recognition device is placed may have an inadvertent affect on recognition accuracy. For example, speech recognition in an aircraft cockpit may be degraded due to engine noise or conflicting voice emanating via aircraft radio communications. Or, consider the placement of such technology in a crowded Military Command Center where its reliability can be affected by background noise from other members located in the nearby work space.

C. APPLICABILITY OF COMPUTER RECOGNITION OF SPEECH

1. Commercial Applications

The first voice input systems to be used by industry were installed in late 1972 and early 1973 [Ref 15]. These early applications included:

-- quality control and inspection

-- automated material handling

-- direct voice input to computers

Their successful implementation was due in large part to recognition accuracies that were greater than or equal to the manual keying accuracies obtained from the same personnel.

In most quality control and inspection processes the inspector's hands and/or eyes are occupied in the inspection task. Through the use of a voice recognition system it is possible to combine the inspector's normal work requirements with the simultaneous entry of all data measured and observed. Owens-Illinois Corporation installed voice data entry equipment in early 1973 for the inspection of color television faceplates. Here was an application where the inspector "had to manipulate, orient, and measure parameters using gauges and meters". The requirement to simultaneously record the measurement data also existed. In this example the operator was able to achieve both tasks at once [Ref. 6: pp. 182-183].

Voice entry has been utilized in recent years to control the movement of materials such as parcels, containers, baggage etc. through distribution and sorting centers. A voice controlled package routing system installed by SS Kresge in November 1974 allowed just one operator to, handle each item, read the label, and speak the destination code for each carton into his/her microphone.

Formerly this had been an operation that required two persons and still resulted in the 'bunching' up of different size packages. Following the installation of voice activated sorting equipment, the bunching problem was eliminated, productivity increased, and sorting errors reduced [Ref. 6: p. 185]

2. Military Applications

These applications may be placed in the general categories of, equipment and process control, field data entry, data management, and cooperative man- machine tasks. A more definitive classification was proposed by Beek et. al. in 1977 [Ref. 16] to include the general areas of Security, Command and Control, Data Transmission and Communication and Processing Distorted Speech. Table I provides a recapitulation of military tasks that could be considered for speech recognition technology.

Of particular interest is the use of speech recognition for Command and Control applications. The term C3, Command, Control, and Communications, refers to an overall system comprised as a minimum of these key elements.

a. Command Authority: The commander provides the central authority, unity of purpose, and the overall concept as to how operations will be conducted to accomplish mission objectives.

TABLE I

MILITARY APPLICATIONS FOR SPEECH RECOGNITION
(From Reference 16)

I. SECURITY

- A. Speaker Verification (authentication)
- B. Speaker Identification (recognition)
- C. Determination of emotional effects (ie. stress)
- D. Recognition of spoken codes
- E. Secure access voice identification
- F. Surveillance of communication channels

II. COMMAND AND CONTROL

- A. System control (ships, aircraft, situation displays, etc.)
- B. Voice operated computer input/output
- C. Data handling and record control
- D. Material handling (mail, baggage, publications)
- E. Remote control (hazardous materials)
- F. Administrative record control

III. DATA TRANSMISSION AND COMMUNICATION

- A. Speech synthesis
- B. Vocoder systems
- C. Bandwidth reduction
- D. Ciphering/coding/scrambling

IV. PROCESSING DISTORTED SPEECH

- A. Diver speech
- B. Astronaut communication
- C. Underwater telephone
- D. Oxygen mask speech
- E. High ´G´ force speech

b. Organization: This element provides the pathways through which the plans, priorities, and directives of the commander are provided to the force and through which information pertaining to the forces can be provided the central authority. These pathways are found at each echelon in the form of command posts, operations centers, or command centers.

c. Communications: This provides the means for transmitting plans, priorities, and orders to elements of the force and the means by which the forces may inform the Commander of their activities and needs.

d. Information: A key element that facilitates control by confronting the Commander with only that information required to support the decision-making process. Information supports both the staff planning and command decision-making process at all levels.

The command centers that will provide the requisite organizational framework, perform several vital functions for the Commander. First, is the capability to communicate securely, and preferably by voice over a wide choice of circuits. Secondly, each command center has the task of integrating information which comes from its supporting elements. A third capability provided by these centers is the processing and display of information. The fourth function, associated with number three, is the quick and

accurate dissemination of information, reports, and directives for the Commander.

We are particularly interested in the function of information processing and dissemination as it provides a suitable application for computer recognition of speech. Command center automation, resulting in more efficient communications, will lead to increased productivity. In its broadest sense, communication is the management of information, and information, not paper, is the chief product of the command center. Our C3 systems that are designed and fielded for these centers, and speech recognition as a component of such, can provide our Commanders the capability to "observe", "decide", "act", and "react" with speed, decisiveness and accuracy.

Navy feasibility studies sponsored by Naval Electronics Command and conducted by Dr G.K. Poock of the Naval Postgraduate School, examined the potential for voice data entry for Command, Control, and Communications. Two voice recognition systems were installed in 1980 at Fleet Headquarters, Commander-in-Chief Pacific (CINCPACFLT) in Hawaii to examine the benefits and limitations of voice input for operation of the Worldwide Military Command and Control Time-Sharing System (WWMCCS TSS) and the Ocean Surveillance Intelligence System (OSIS) [Ref. 17: p. 34].

Poock has also demonstrated that using voice input to exercise a typical scenario on the ARPANET, an experimental network since 1969 employing packet switching technology and connecting over 150 host computers, was significantly faster and more accurate than entering the commands manually [Ref 8]. Twenty-four subjects followed a fixed scenario of instructions where they accessed the ARPANET, logged into different host computers, read messages, sent messages, read files, transferred files between host computers, deleted files and interconnected host computers. Simulated command centers operating on this network include the Naval Postgraduate School (Monterey, California), Naval Ocean Systems Center (San Diego, California) and CINCPACFLT (Hawaii).

Automatic speech recognition has also been found to have considerable potential for imagery interpretation and intelligence report generation [Ref. 17: p. 49]. A significant amount of research has been performed for the Defense Mapping Agency (DMA) for such applications as voice data entry for the processing of Digital Landmass System (DLMS) data, preparation of Flight Information Publication (FLIP) data and ocean-depth measurements for digitized cartographic applications. In all these applications the environment is such that the operator's hands are busy and frequently involve the use of stereo optics and other special devices. Voice has been shown experimentally to be

faster, easier, and a less fatiquing mode of data entry than historically more conventional means [Ref. 17: p. 37]. More recently, the feasibility and advantages of voice input technology were described for use in the COINS Network Control Center (CNCC). The Community On Line Intelligence System interconnects on-line information storage and retrieval systems located at a number of locations within the United States intelligence community [Ref. 18].

III. HUMAN FACTORS IN SPEECH RECOGNITION

A. DEFINITION AND PURPOSE

Human factors is concerned with improving the productivity of the user by taking into account human characteristics in the design of a system. As described by Huchingson [Ref. 19: p. 4],

> The term "human factors" is more comprehensive, covering all biomedical and psychosocial considerations applying to man in the system. It includes not only human engineering, but also life support, personnel selection and training, training equipment, job performance aids, and performance measurement and evaluation.

The people referred to in this definition are those who typically operate, maintain or service the system. They are those who will interact with the system's design. When the focus is on a broader interpretation it's appropriate to speak of a Human Factors Subsystem or Personnel Subsystem as was described earlier.

Human factors engineering deals principally with the many factors involved in the design of a new system - from hardware to personnel. For our efforts in this analysis, the current technology has been determined to be acceptable and, experimentally as well as operationally reliable for its use in a Command and Control environment. Now, user variability is to be investigated further in terms of how it affects recognition accuracy.

Since energy in a speech signal is usually displayed in terms of frequency, intensity and time, it would seem plausible that each word should have a unique acoustic wave pattern and, if so, word recognition would be a simple matter of the voice recognition system scanning the pattern, comparing the simple pattern with a data bank of reference word patterns, and deciding which word was spoken. Unfortunately, human variability messes up this uniquely simplistic approach. Our purpose then is to discuss the human as a component in a complex system designed by humans and to note the fundamental advantages and limitations of the human in relation to an automated voice recognition system.

B. FACTORS AFFECTING RECOGNITION ACCURACY

1. General

Limitation of vocabularies to 100 words have resulted in identification accuracies of between 98% - 99% in a controlled laboratory environment. In an operational or field setting recognition accuracies have been reported as low as 50% [Ref. 20: p. 636]. Various factors noted for interfering with successful identification have included background noise, inconsistent microphone placement, insufficient training, inconsistent speaking style, and the lack of user cooperation. Lea in a paper titled "What Causes Speech Recognizers to Make Mistakes?" [Ref. 21] calls

for the determination of those factors that influence recognition accuracy rather than the repeated assessment of transitory devices. Table 2 summarizes the four 'dimensions of difficulty' Dr Lea has proposed. What needs to be accomplished is the characterization of the relative effects of changes along each of these four dimensions, or more simply stated, find the factors influencing the accuracy of machines that recognize speech.

Because there are so many variables involved that affect recognition accuracy, the list in Table 2 may be reorganized in a "communication-theoretic" framework. This framework models the speech recognition error rate as a function of seven complex sets of factors [Ref. 5: pp. 69-93] that include:

-- Task Factors
-- Human Factors
-- Language Factors
-- Channel and Environmental Factors
-- Algorithmic Factors
-- Performance Factors
-- Response Factors

It is the set of Human Factors that this experiment and analysis is principally concerned with, for it is this stage of the model that has a major impact on speaker

TABLE II
DIMENSIONS OF DIFFICULTY FOR SPEECH RECOGNITION
(From Reference 5)

TASK AND PERFORMANCE REQUIREMENTS	1. Form of speech to be recognized 2. Accuracy requirements 3. Required throughput rates 4. Type of device necessary
HUMAN VARIABILITY	1. Sex 2. Dialect 3. Vocal tract size 4. Vocal cord characteristics 5. Pronunciation habits of speaker 6. Physical state 7. Psychological state 8. Workload 9. Cooperativeness 10. Time of day/week 11. Time since training 12. Number of training samples/word 13. Rate of talking
LANGUAGE DIFFICULTIES	1. Size of active subvocabulary 2. Word length 3. Word sound structure 4. Confusability 5. Language spoken 6. Syntactic, semantic, and pragmatic constraints 7. Enhanceability 8. Stress Pattern 9. Intonational variability 10. Rhythm and timing variability
ACOUSTIC DIFFICULTIES	1. Noise level 2. Type(s) of noise 3. Bandwidth 4. Spectral distortions 5. Transducer characteristics 6. Placement of the transducer 7. Amplitude 8. Vibration 9. Acceleration

variability. This set of human factors can be further subdivided [Ref. 21: p. 2] in order to monitor their influence on recognition error rates. A few of these are listed below:

- -- Speaker Experience
- -- Training Method
- -- Sex of the Speaker
- -- Physical Dimensions of the Speaker
- -- Geographic Origin of the Speaker
- -- Speaker Dialect
- -- Physical State of the Speaker
- -- Psychological State of the Speaker
- -- Speaker Cooperativeness
- -- Time of Day or Week

Because different speakers may demonstrate widely varying methods of pronouncing words or phrases, the above listed factors may be further separated into two categories; those occurring between speakers and those affecting each individual speaker. First, some of the differences between speakers that induce variability will be briefly examined and then the variabilities apparent within each speaker that can affect recognition accuracy will be discussed.

2. Differences Between Speakers

Speaker Experience: This factor can take on a two-fold meaning when looking at it as a source of variability.

First is the experience of using voice recognition equipment. Experienced voice recognition users should be expected to have a higher and more reliable recognition accuracy than those who are ´naive´ to the technology. These experienced users are comfortable using the equipment, less likely to be intimidated by the system, and are familiar with its performance capabilities from previous usage. The other meaning of speaker experience has to do with job skill. Can a user who operates in a microphone environment on a daily or regular basis, such as an Air Traffic Controller or a Pilot, be expected to have better recognition rates than those who have never spoken into a microphone? A data processor who works regularly in an environment demanding precise data entry by keyboard might have the type of experience or skill factor that would provide an edge over a prospective user possessing only basic typing skills. This type of experience overlaps slightly with speaker cooperativeness and will be elaborated upon later.

Method of Training: The ideal form of voice interaction would be for a user to pick up the microphone, speak commands the machine can understand, and for the appropriate response to take place. Naturally, this is the goal of speaker independent systems, but since humans all speak differently and our form of speech recognizer is discrete, we are mandated to provide the machine some

information about how we speak each word intended for our desired vocabulary (ie. Training). The method by which the machine is trained by the user will in large part dictate subsequent recognition accuracy. If the user is closely supervised and made to carefully speak the particular vocabulary then we should be able to expect higher recognition rates as opposed to the user who is given cursory instructions on the use of the equipment and allowed to go on independent of further supervision during the training mode. An adjunct of training method is the number of training ´samples´ or pronunciation pattern. It is difficult to achieve accurate speech recognition when the number of training passes per word is small or smaller than manufacturer specifications [Ref. 22]. Using identical equipment, it would still be reasonable to anticipate some speakers, having had a lesser amount of training samples per word, having more success than others who have had more samples per word.

Sex: Male voices have lower frequencies than females and a more detailed spectral structure results from the lower pitch of their voices. This detailed structure is more indicative of the vocal mechanism and of the intended vowels and consonants spoken. Male voices tend to fare better with recognizers employing frequency domain analysis while female voices tend to have greater success with machines using time domain analysis [Ref. 5]. A recent

comparison was conducted [Ref. 22] which revealed no statistically significant difference between the sexes. Although not a primary objective of the thesis, it remains a source of variability that merits some measure of analysis.

Speaker Dialect: Dialects not only affect the specific sound produced for each vowel or consonant type, but also exhibit different dynamics of speech production. For example, Southerners have their readily identifiable drawl, whereas a New Yorker will tend to say "Toid" rather than "Third" and residents of Cambridge, Massachusetts can be heard to talk about "Hahvahd" instead of "Harvard".

Physical Dimensions: Throughout the literature on speech recognition one will see speaker variability attributed to a variety of factors, none of which include the physical dimensions of the speaker. An examination of the recognition accuracy for a selected sample population based on physical dimensions would provide an interesting insight into the ramifications of such a factor as a component within a personnel selection subsystem. In other words, what effect, if any will height and weight have on recognition accuracy?

Geographic Origin: This particular factor is multidimensional consisting of several sub-factors which require careful examination:

-- Place of birth

-- Geographic area of upbringing

-- Ethnic background

-- Religious preference

The above may impose ideosyncratic or social differences in habits which can produce variations in sound and subsequently in pronunciation. These sub-factors all contribute a measure of variety that can presumably affect recognition accuracy.

3. Differences Within Speakers

Physical State: The present physical state of a user of voice recognition equipment can precipitate variability in his or her voice. For example, a cold, some form of pathological condition, fatigue etc. can alter the speaker's voice. The individual's voice quality could be different based on physical conditioning. Is the user who works out regularly and stays in excellent physical condition more likely to show higher recognition rates than one who rarely exercises, smokes regularly and generally is not in the best of health?

Psychological State: Spielberger [Ref. 23: p. 29] defines transitory or state anxiety as a complex, unique emotional condition that can vary in intensity and fluctuate over time. State anxiety may be thought of as consisting of unpleasant, consciously perceived feelings of tension and apprehension with an accompanying activation or arousal of the autonomic nervous system. The concept of trait anxiety refers to the relatively stable individual differences in

anxiety proneness. It may also be a reflection on the frequency and intensity with which state anxiety has been previously manifested and the probability that such anxiety will occur in the future [Ref. 23: p. 39]. The fact that physiological functioning is affected during periods of anxiety is easily apparent. The degree to which speakers deal with a state or trait anxiety may well be a significant variable of consideration in the examination of error rates of voice recognition systems.

Speaker Cooperativeness: How enthusiastic and/or willing a speaker is toward the use of voice recognition equipment could induce speaker variability and hence subsequent recognition accuracy. In a military environment where many job positions are of a non-voluntary variety, it is conceivable to expect the selection of voice recognition users who are told to operate the equipment regardless of their personal preferences. If the user distrusts the technology or prefers manual entry, and, is still required to use voice, we have developed a non-cooperative user. A non-cooperative user is therefore, one who is consciously trying to undermine the successful operation of the machine. The cooperative user is one who is willing to help the machine by saying precisely what the machine wants and pronouncing it in a clear and consistent manner. There is a certain grey area surrounding this factor with the presence of users who, although not consciously trying to confuse the

device, are not fully committed to "helping the machine" to recognize the correct utterances.

Time of Day/Week: Each person's speech is variable depending upon time of day, changing from morning to evening and even changing progressively over a period of time [Ref. 5]. An examination of recognition performance over extended periods of time [Ref. 24: p. 1] showed a statistically stable performance over time (21 weeks) with no serious degradation occurring as time elapsed. Nevertheless a user who has a gap in time between training and operational use may forget any special ways he/she trained the machine. How much of a gap is tolerable is a subject for future research.

4. Miscellaneous Factors

Some additional human factors that have been proposed [Ref. 5] deserve a brief description. They have been relegated to a separate section because, for one reason or another, lack of equipment, current technical skills, lack of measurable quantitative data etc. experimental examination at the present time has been precluded. These factors include:

- -- Form of speech
- -- Speaker dependence
- -- Rate of speech
- -- Vocal tract size
- -- Speaker's glottal spectrum

Form of speech refers to the type of voice recognition system to be used, isolated or continuous. Continuous systems, being a quantum step above isolated in terms of complexity, bring about a greater opportunity for speaker variability to manifest itself. Such things as detection of word boundaries, slurring of speech (ie. "dija" vs "did you"), and prosodic characteristics could seriously affect recognition accuracy because of these types of complications which a continuous speech recognition system introduces.

A speaker independent system negates the requirement for training and thus variability between speakers becomes a more critical factor for independent systems to contend with. Independent recognizer performance will have to be tailored to accommodate an unlimited number of potential speakers and their associated variability.

The faster a person speaks the more likely that the expected pronunciation will be altered due to slurring, deleted syllables, etc.. If a machine is trained to one form of pronunciation and at one particular rate of speech, a differing rate in an application mode, will cause an increase in recognition difficulty. With an isolated word recognizer to be used in the experimentation, requiring a minimum of 100 msec pause between utterances, and utterances not exceeding 2.0 seconds in duration, this particular factor was not considered essential to the overall analysis.

It is rather, an important factor in terms of continuous recognition systems.

The size of the vocal tract will produce changes in the formants of the speech signal; the smaller the vocal tract the higher the formants. This can have an impact on, for example, transmission through limited bandwidth channels. Vocal cord characteristics also produce interspeaker variability such as pitch or "resonant" quality of the voice. Speakers with more "resonant" voices that project well, will be easier for recognizers to handle [Ref. 5: p. 78].

IV. DESCRIPTION OF THE EXPERIMENT

A. OBJECTIVES AND CONSTRAINTS

1. Objectives

As noted earlier, our overall objective was to examine the human as a component in a complex system. In narrower terms, this experimentation attempts to assess the affect of differing occupational, operational, personal, physiological, and psychological characteristics of a user, on the accuracy with which a currently available voice recognition system will correctly interpret spoken utterances. Subsequently, our discussion will address the occurrence, if any, of existing quantitative parameters that would enable us to differentiate between effective and non-effective users of voice recognition systems.

The following specific characteristics are examined in this thesis. Many of the individual characteristics, or human factors, are self-explanatory while others are provided with a brief explanation and/or rationale for selection.

a. Occupational Characteristics

This set of parameters examines the possible effect on recognition accuracy due to differences inherent in a user's occupational skill or job (military or civilian) background. Specific characteristics include:

-- Job function: Comparison of recognition rates between microphone experienced users (ie. pilots, air traffic controllers) and non-experienced users.

-- Branch of service: A factor with possible consequences pertaining to its use in personnel selection criteria.

-- Job satisfaction: A subjective evaluation by the user as to his/her job satisfaction in their current duty assignment and their satisfaction within the Armed Services.

-- Previous computer experience: Computer experienced personnel (ie. Data Processors) are expected to have a better appreciation for the advantages of voice input and thus, be more conscious of their efforts and positively motivated for higher recognition accuracy.

-- Foreign language competency: Frequently military and civilian members associated with DOD are required to possess the capability to fluently speak a foreign language. This ability is another factor that could affect one´s speech.

b. Operational Characteristics

This set of parameters examines the possible effect on recognition accuracy due to factors surrounding the operational use of voice recognition equipment. Specific characteristics include:

-- Training method: Analysis of recognition rates for those users who are supervised during the training mode compared to those who are allowed to train the equipment individually.

-- Time of day and week: A determination of whether the time frame in which a speaker trains the recognizer will have any subsequent affect on recognition accuracy.

-- Equipment experience: Comparison of recognition rates between experienced users of voice recognition equipment and those who have never used the equipment before ('naive' users).

-- Ease of use: The operational simplicity of the equipment could affect a speaker's performance. For example, a speaker who considers the recognizer as a complex and operationally difficult device will be less likely to devote his or her maximum effort to their performance.

c. Personal Characteristics

The following are various characteristics considered to have a possible effect on an individual's speech patterns, and hence, affect the recognition accuracy of a voice system. These parameters include:

-- Race

-- Marital status and family size: A correlate of

psychological state and, although equally likely to be included as a psychological characteristic, it is considered here as a criterion for personnel selection. Family size refers to the number of offspring the user has as opposed to the size family in which one was raised.

- Religious preference/Ethnic background
- Accent or dialect
- Place of birth/geographic origin
- Level of education
- Socioeconomic class: similar in nature to the characteristic of marital status but is considered for its merit in selection of personnel than for its affect on individual speech patterns.
- Dental or orthodontal care: Braces, corrections for improper bite, or major oral surgery, are considered for their implication on the speech patterns of those individuals and the resultant error rate.

d. Physiological Characteristics

These characteristics are also considered to have an affect on speech and as a result are factors of interest when examining recognition accuracy and speaker variability. These parameters include:

- Height
- Weight

-- Age

-- Physical condition: A subjective evaluation by the user of his/her current physical condition.

-- Rate of airflow: Measurement of ventilatory function to provide a diagnosis of condition affecting voice. This measurement can also be used as an indication of possible airway obstruction.

-- Vital capacity: The maximum amount of volume of air which can be exhaled following maximum inhalation. This measure provides an estimate of the amount of air potentially available for the production of phonation.

-- Speech training: Examines whether formal speech or voice training affects recognition accuracy.

e. Psychological Characteristics

The current psychological state of a user, their cooperativeness, and their personal attitudes toward automation and voice all contribute toward the overall affect on recognition accuracy. The particular parameters investigated include:

-- Psychological anxiety

-- Speaker cooperativeness

-- Affect of errors on subsequent performance

-- Attitudes toward voice recognition equipment as a time saving job aid

-- Attitudes towards computers and data automation.

In effect, items 4-6, are related to speaker cooperativeness in that how a user feels about computers and voice recognition could impact on their willingness to reliably support the use of voice recognition equipment.

2. Constraints

Accomplishment of test objectives were constrained within the research facilities of the Naval Postgraduate School. In the interest of time, experimentation was limited to five weeks.

Because voice production is an extremely complex event in which auditory, acoustic, and aerodynamic events are produced by the interaction of physiological mechanisms, it would be beneficial if we could measure as many vocal parameters as possible in order to achieve a complete and accurate picture of voice production, its associated variability among speakers, and its correlate to voice recognition accuracy. Lack of equipment, time, and/or expertise precluded examination of such factors as:

-- Glottal waveform
-- Transfer function of the vocal tract
-- Sound-pressure level
-- Maximum duration of sustained phonation
-- Maximum frequency levels
-- Modal frequency level

B. SUBJECTS

Forty-four subjects participated in the experiment on a volunteer basis. The group was composed of 25 military officers, 17 military enlisted, and 2 civilians. The military officers representing the Army, Air Force and Navy consisted of 21 males and 4 females while the enlisted personnel representing the Army and Navy consisted of 11 males and 6 females. The civilians included a professor from the NPS Oceanography Department and an employee of the Defense Manpower Data Center (DMDC) in Monterey. The rank or grade of the military subjects ranged from O-2 to O-4 for the commissioned officers, CW2 to CW3 for the Warrant Officers, and E3 to E7 for the enlisted personnel. The subjects ages ranged from 20 to 47, with an average age of 30.

It was desired that the speakers selected for the test be representative of the population for which the recognizer is to be used, in our case a Command and Control environment and in particular, a military command center. Subjects taking part in the experiment were representative of this environment as shown by the grade distribution and types of military occupational specialties, although some of these specialties are not readily apparent in current job description (ie. Medical NCO).

Twenty-five of the subjects were from Fort Ord and included a variety of backgrounds such as pilots, air

traffic controllers, signal officers, signal non-commissioned officers (NCO's), and infantry platoon sergeants. Five of the subjects were data processors; 2 from the Fleet Numerical Oceanographic Center in Monterey and 3 from administrative offices of the Naval School. Twelve subjects were students at NPS and enrolled in the Command, Control, and Communications (C3) curricula. A wide diversity in their backgrounds is illustrated by previous job categories such as aviation, communications, systems programming, communications maintenance, command and staff, and nuclear engineering.

Twelve of the subjects had experience using voice recognition equipment, having participated in previous voice experimentation [Ref. 9]. A summary of subject characteristics is provided in Table III.

C. EQUIPMENT

1. Voice Recognition System

A Threshold Technology Inc., Model T-600 voice recognition system was used to represent a commercially available, state-of-the art recognizer; one which has been well documented as to its reliable recognition accuracy. The T-600 is a speaker dependent, isolated word, speech recognition device which automatically recognizes spoken words and phrases. These words and phrases (utterances) may be as brief as 0.1 second but will usually range from 0.25

TABLE III

SUBJECT CHARACTERISTICS

SEX	SERVICE	LOCATION	VOICE
Male: 34	Army: 27	Ft Ord: 25	Experienced Users: 12
Female: 10	Navy: 6	NPS: 16	Naive Users: 32
	Air Force: 7	FNCC: 2	
		DMDC: 1	

RANK	OCCUPATIONAL BACKGROUNDS	
O-4: 6	Pilots: 2	Air Traffic Controllers: 5
O-3: 9	Data Processors: 5	Supply Officer: 2
O-2: 5	Medical Officer: 1	Medical NCO: 1
CW3: 2	Signal Officer: 3	Signal NCO: 3
CW2: 3	Finance Officer: 1	Engineer NCO: 1
E-7: 5	Operations Officer: 1	Professor: 1
E-6: 4	Computer Systems Manager: 1	
E-5: 7	Graduate Students: 12 (which include)	
E-3: 1	Pilots: 3	
CIV: 2	Communications Officer: 2 Communications Maintenance Officer: 2 Systems Programmer: 1 WWMCCS Programmer: 1 Submarine Nuclear Engineer: 1 Infantry Unit Commander: 1 AUTODIN Supervisor: 1	

to 1.0 seconds and must be separated by very short pauses of .1 second or more. The terminal allows a user to begin an utterance before it has completed processing the previous one, but in this experimentation rate of speech was controlled by use of the READY indicator light located on the tape cartridge unit. This light indicates when the terminal is ready to accept the next utterance in both the training and recognition modes [Ref. 25].

The Threshold 600 in its standard configuration is composed of the following four elements:

-- Terminal consisting of:
    - analog speech preprocessor
    - LSI-11 microcomputer
    - digital RS-232 input/ouput interface
-- Standard CRT/Keyboard Display Terminal
-- Remote Voice Input Unit (Microphone preamplifier)
-- Tape Cartridge Unit

The terminal, CRT display, microphone preamplifier, and tape cartridge unit were table mounted (Figure 3) within an acoustic sound reduction booth (Figure 4). A conventional SHURE model SM-10 "boom" microphone, supplied as standard equipment with the T-600 was used. The microphone possesses a special noise cancelling design which allows the T-600 to perform accurately despite most extraneous background noises (Figure 5).

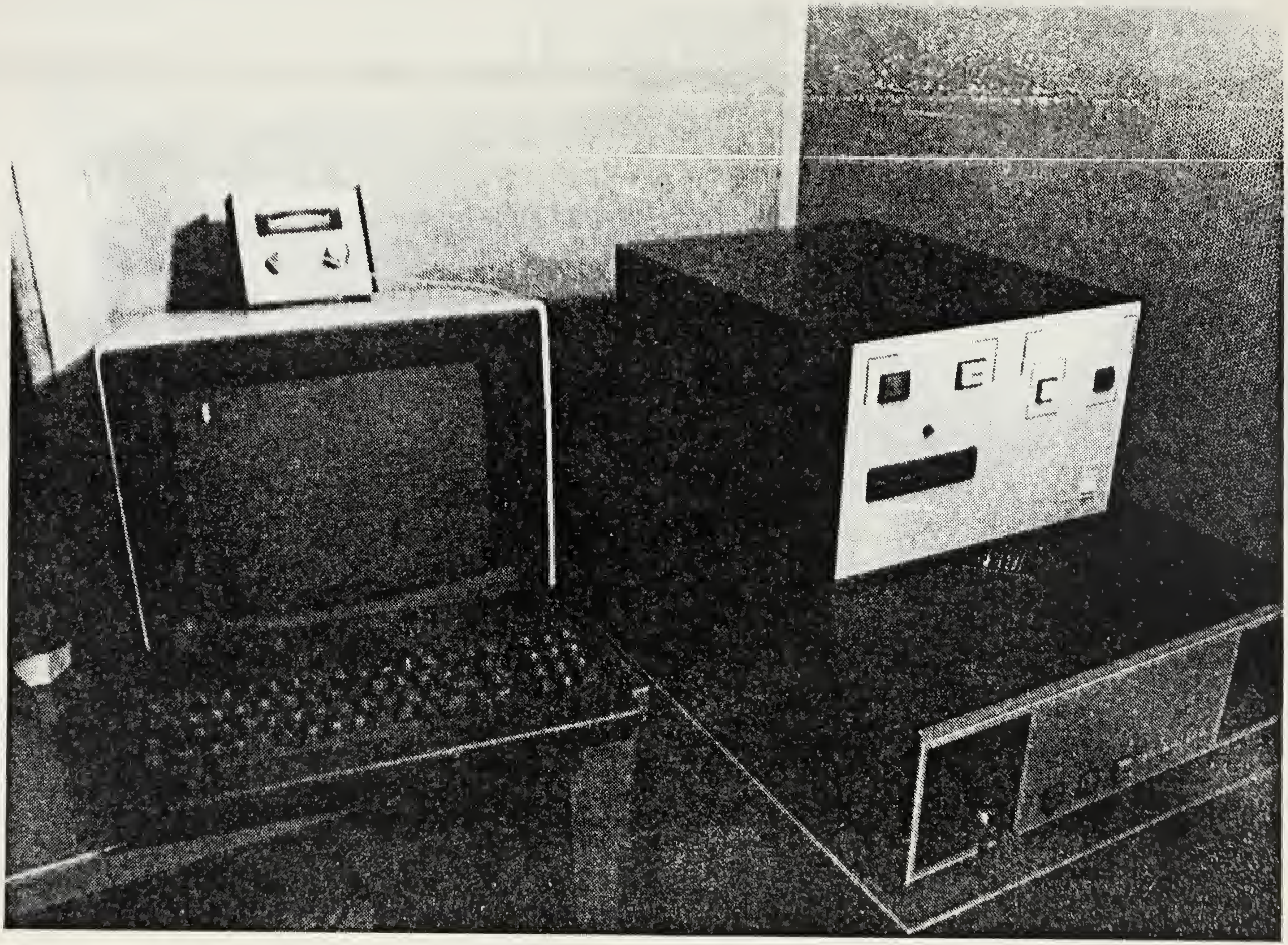

Figure 2. T-600 Speech Recognition Equipment

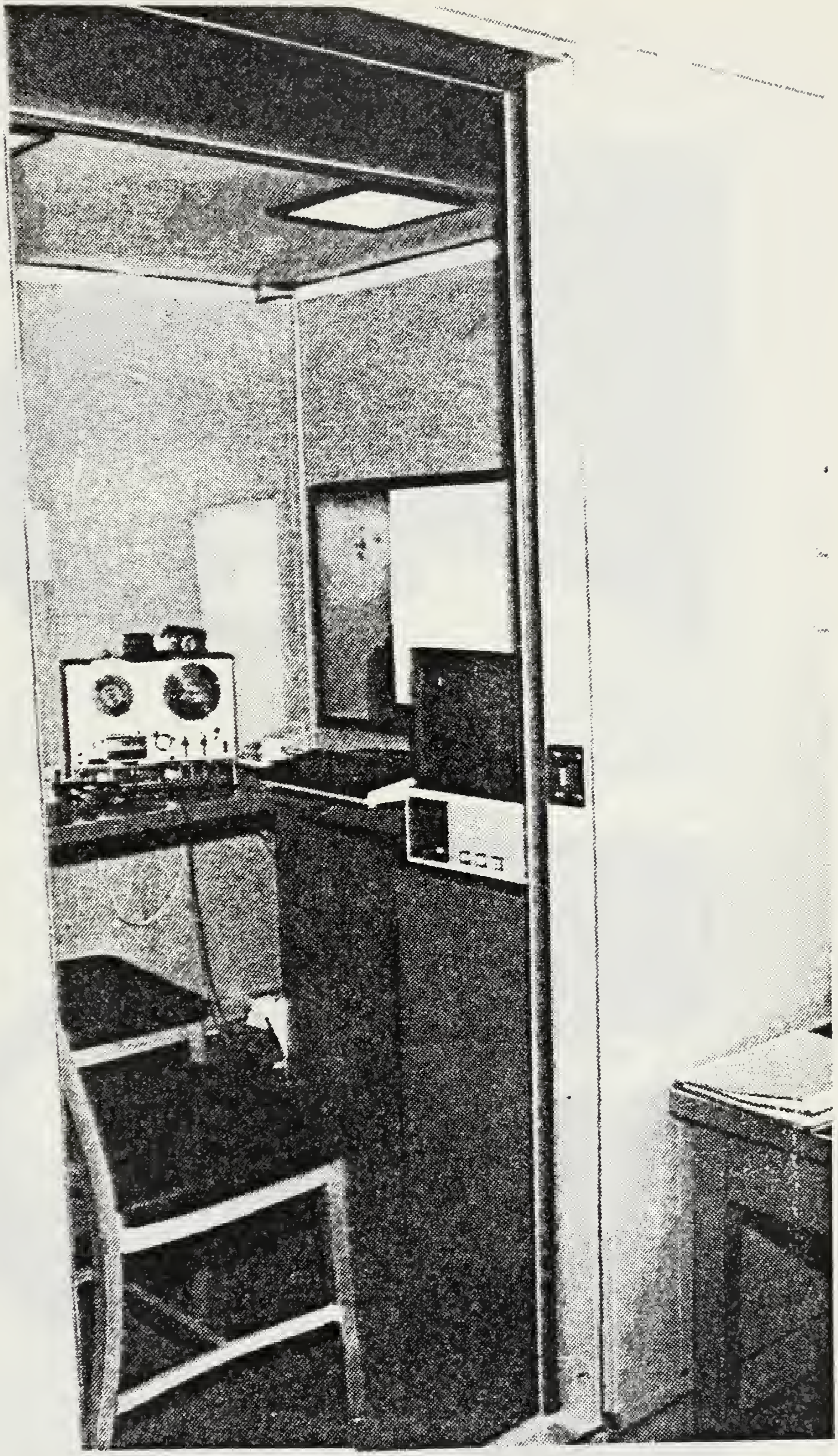

Figure 4. Accoustic Sound Reduction Chamber

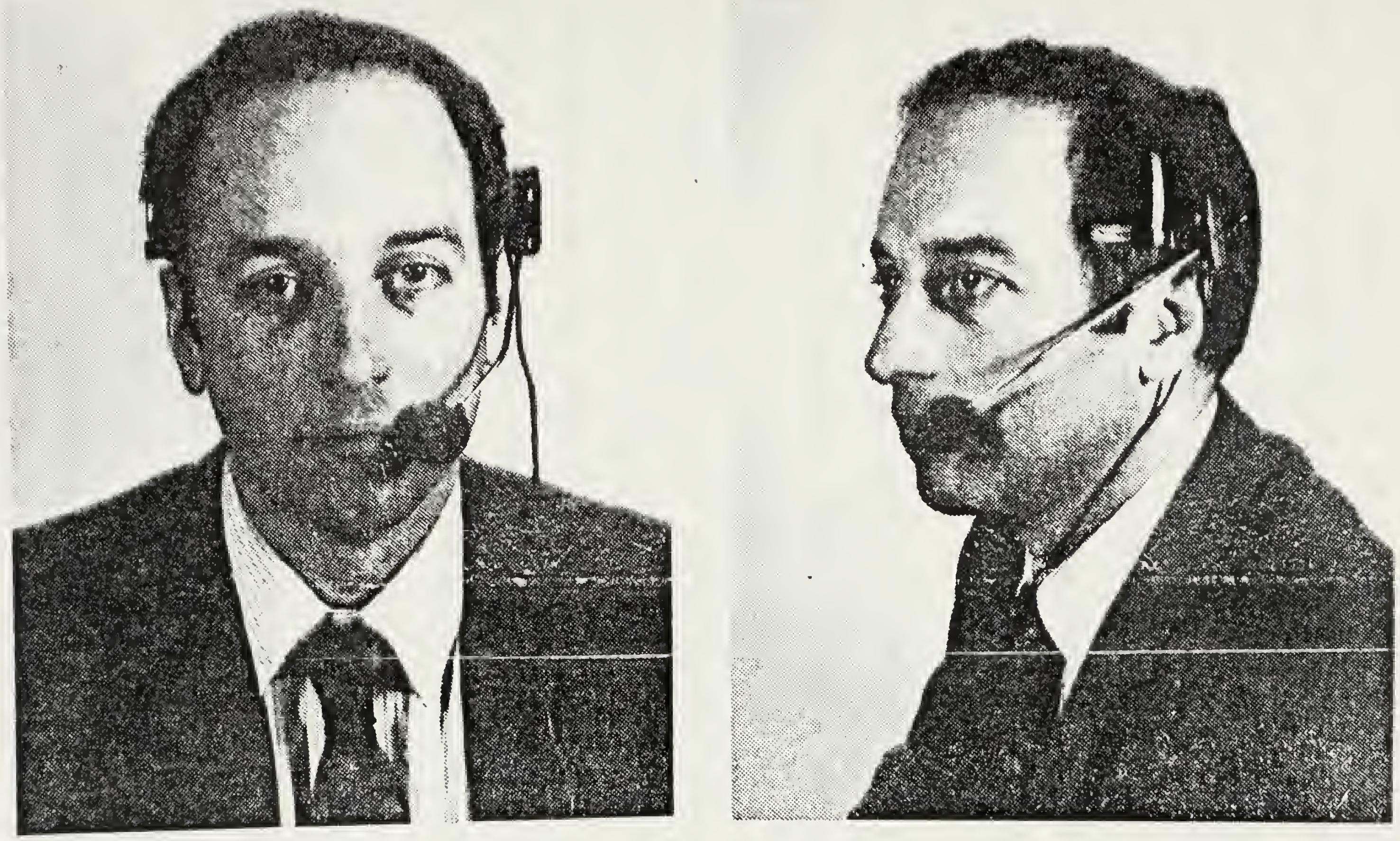

Figure 5. Placement of the SHURE SM-10 Microphone

The speech preprocessor accepts the speech signal input from the microphone preamplifier and passes it through a spectral analyzer for word boundary detection. The feature extractor monitors for 32 phonetically-relevant features, and converts these to digital signals. Words are detected from occurrences of low energy. A minimum pause of 0.1 second must occur to prevent confusion between words. Any breathing noise at the end of the word is removed. The remaining speech is divided into 16 fixed time segments, and features are reconstructed onto the normalized 16 segment time base.

The microcomputer does a comparison of input signals against stored reference patterns. Each word is represented by 512 (16 x 32) bits of information. The closest fit between an incoming template and the alternative stored training template is found, and that ´closest´ word is declared the word identity, unless the score is so low that no decision can be made and the utterance is rejected outright. The vocabulary reference patterns are established by the subject ´training´ the recognizer. This is accomplished by the subject making a set number of repetitions of the various vocabulary utterances.

Once a match is found, the appropriate character(s) are sent via the output interface to the CRT to indicate to the user which utterance was recognized. These terminal matches are further categorized as misrecognitions, where

the terminal's 'closest' match to the reference vocabulary was not precisely the same utterance spoken, or recognitions, in which the utterance spoken is exactly recognized and so reflected in the CRT output. Rejection of an utterance is a third category and is indicated by an audible 'beep'.

The remote voice input unit allows components to be remotely located up to 2000 feet from the terminal processor and provides the means to adjust the volume (amplification) of the amplifier to accommodate the normal speaking voice of each particular subject.

The tape cartridge unit is a digital tape recorder used to store and recall application data and an individual subject's vocabulary reference patterns. Once the data cartridge is recorded it contains all the information necessary to initialize the Threshold 600 terminal for each subject. The T-600 is capable of storing a 256 word vocabulary which may be recorded or loaded in a few minutes using the tape unit.

2. Spirometer

A recording spirometer, Figure 6, a type of gasometer, was used for measuring and recording vital capacity. It consists of a metal tank containing a movable piston with a water seal, air input line, exhaust valve for resetting, ink stylus, and revolving cylinder for mounting chart paper calibrated in cubic centimeters.

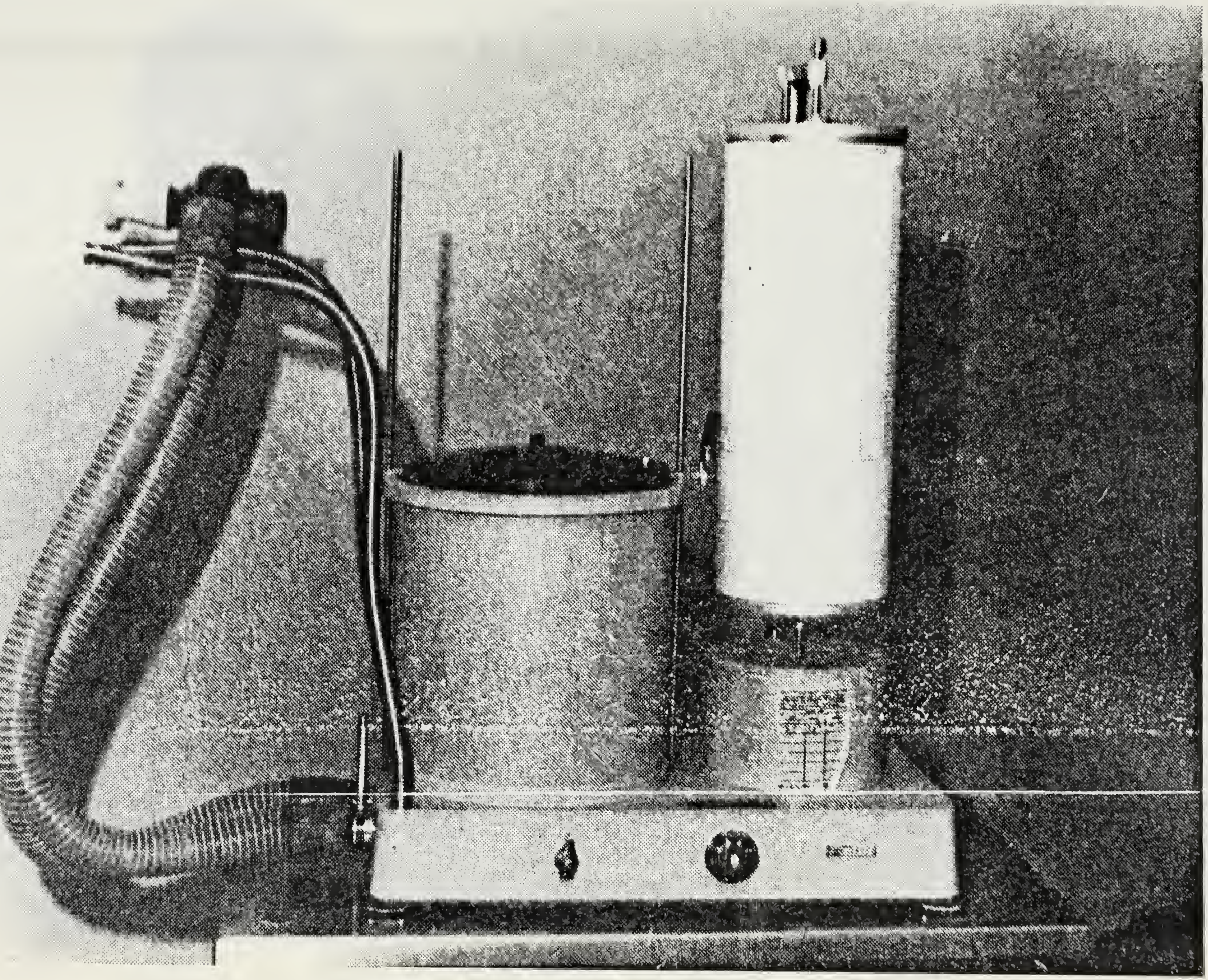

Figure 6. Recording Spirometer

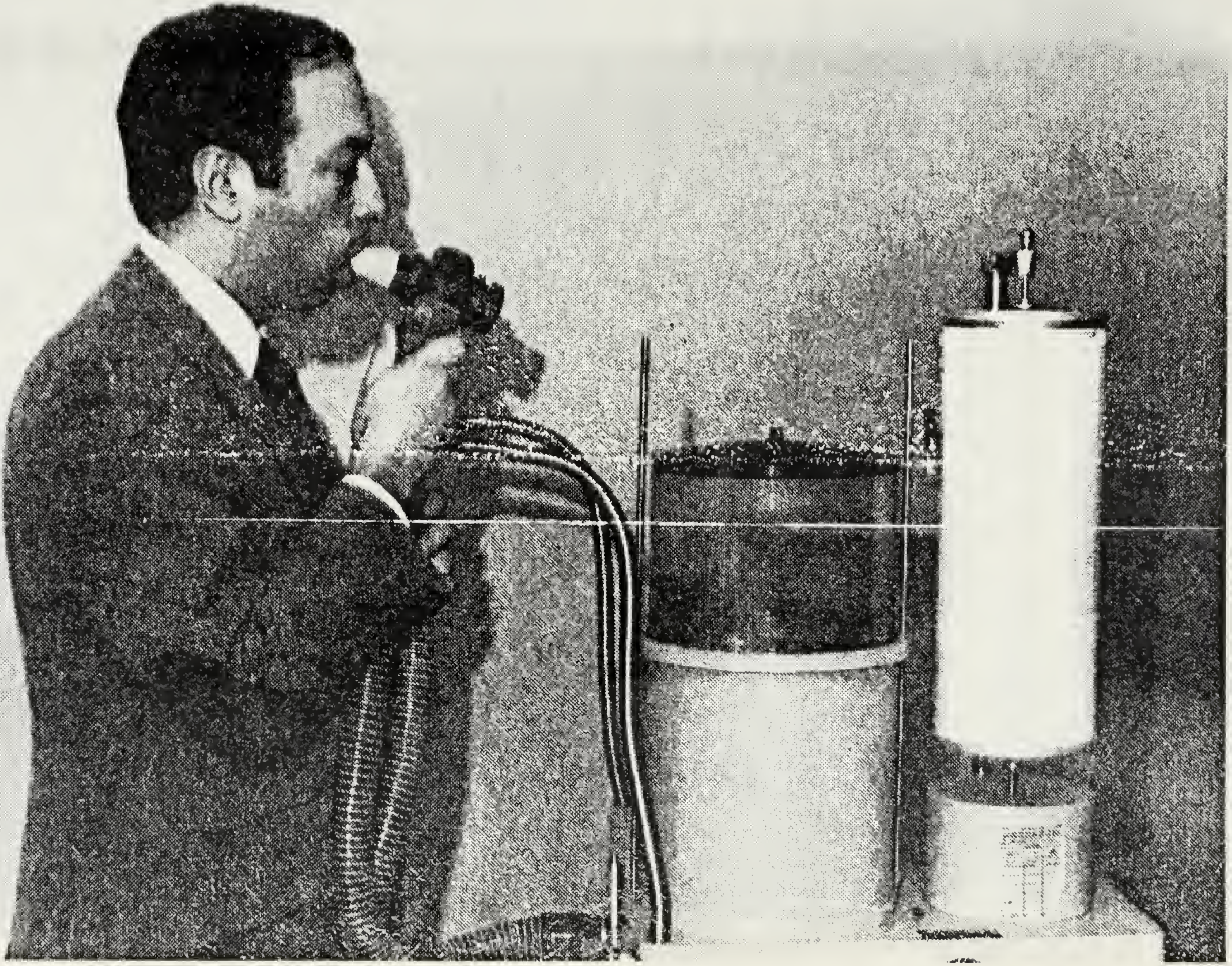

Figure 7. Use of a Recording Spirometer to Measure and Record Vital Capacity

As the subject breathes into the mouthpiece, Figure 7, air replaces water in the inner piston, which rises by an amount proportional to the exhaled air. The subject, once fitted with the mouthpiece, is given instructions to inhale to the greatest extent possible and then exhale all the air. This procedure was repeated three times and the average vital capacity used for analysis purposes.

3. Peak Flow Meter

The Wright Peak Flow Meter was used to measure the maximum air flow rate in a single forced expiration. The instrument, Figure 8, consists of a pivoted vane, the rotation of which is opposed by resistance of a spring. The plastic mouthpiece fits into the radial inlet which leads to the vane. Attached to the vane is a spindle and pointer. The forced expiration causes the vane and pointer to rotate until the maximum attainable flow has been reached. Once reached, the pointer is held in position by a ratchet until released by a reset button on the back of the device. The scale is graduated in liters per minute in 5 liters/minute divisions over a range of 60 to 1000 liters/minute.

Procedurally, the subject stands and holds the meter in a vertical plane as depicted in Figure 9. He/she then takes as deep a breath as possible, places the mouthpiece in the mouth, grips it tightly with the teeth, and seals it with his/her lips. The subject blows out as hard as possible in a short, sharp expulsion of air. This procedure

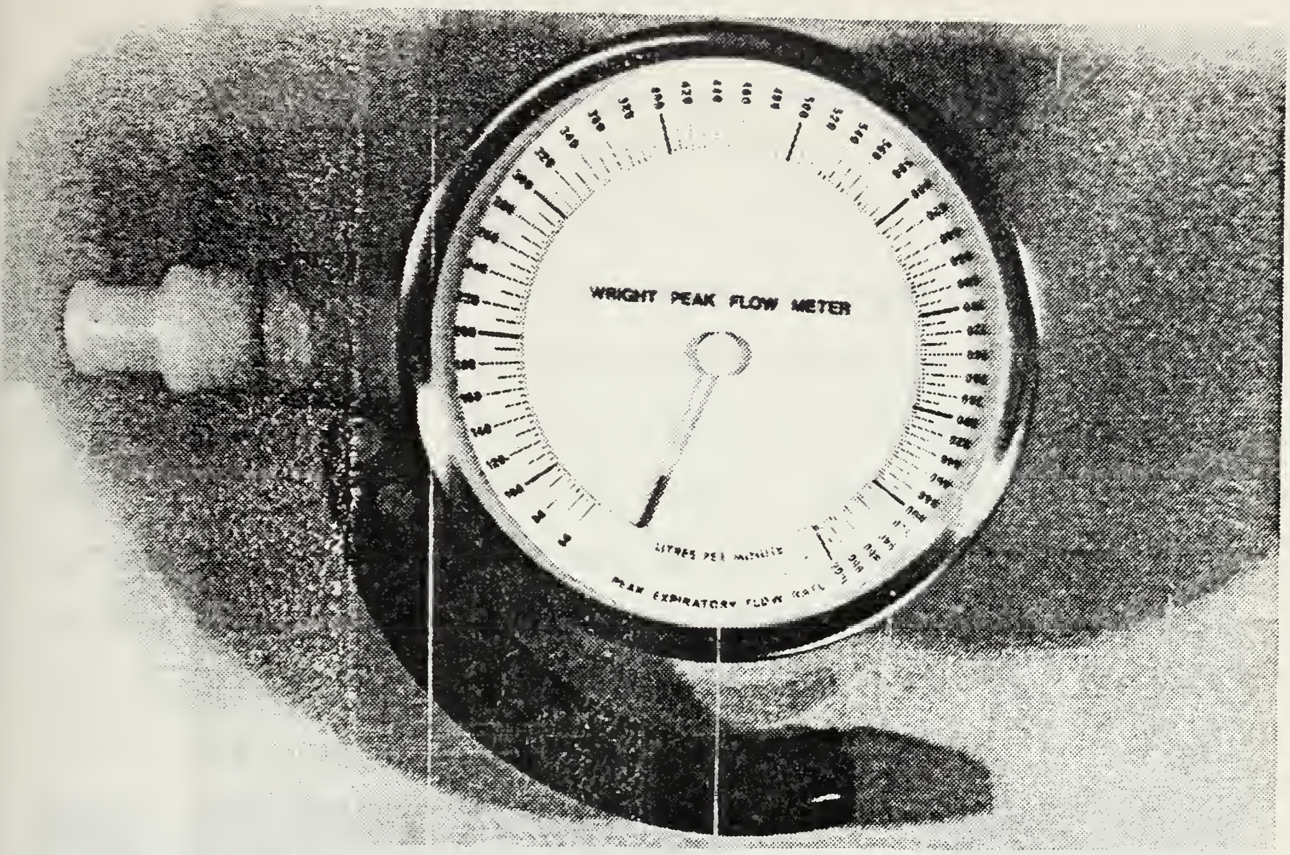

Figure 8. The Wright Peak Flow Meter

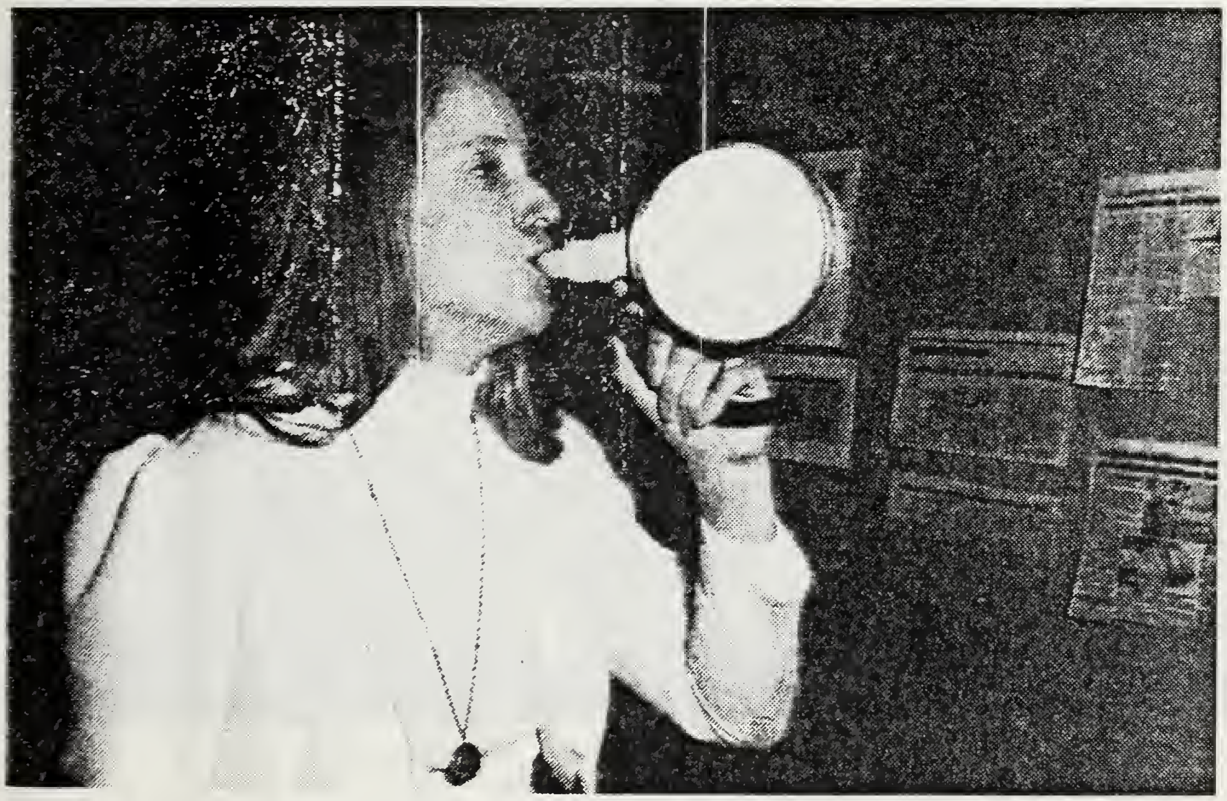

Figure 9. Measurement of Speakers' Rate of Air Flow

was performed three times with the average noted as the appropriate peak expiratory flow.

4. Tape Recorder

An AKAI 4000 DS Mk-II magnetic tape recorder was used for the recording, storage, and reproduction of speech sounds (Figure 10). The device is a typical analog magnetic tape recorder consisting of three basic parts. These include the electronics of the system, the head assembly, and the tape transport. These components take a phenomenon, such as the speech sound, that changes in time and records it as a continuous event.

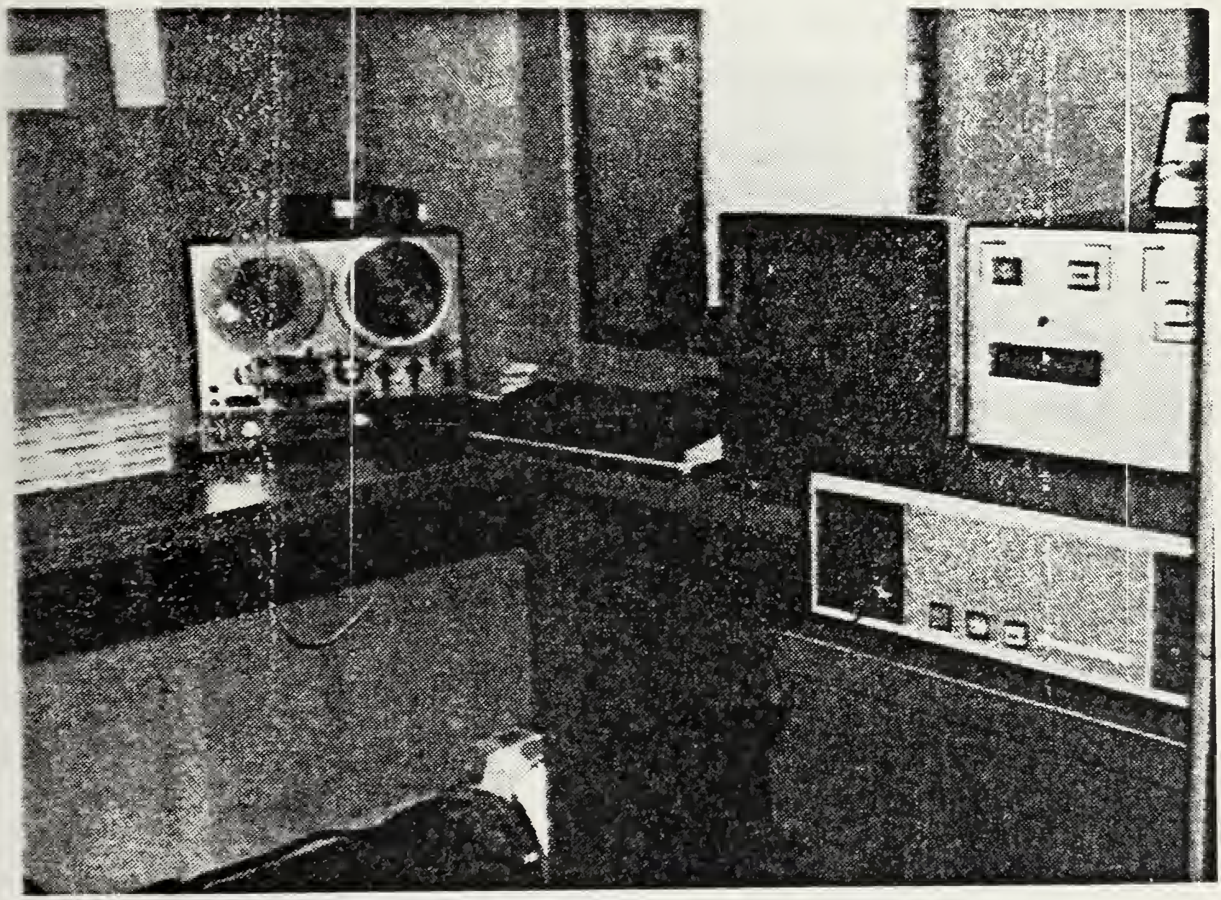

Figure 10. AKAI Tape Recorder

Tapes were recorded for all 44 subjects during their participation in the experiment. Subject to availability of analytical software at NPS, further acoustical analysis could be conducted on speaker variability that might substantiate and support statistical conclusions.

D. INSTRUMENTATION

Three questionnaires were used to elicit the evaluations, judgement, comparisons, attitudes, and background history of the subjects participating in the experimentation. The first two questionnaires were designed [Ref. 26] to provide the necessary information to delineate subjects into various groups representing those human factors discussed earlier. The third questionnaire was used to measure state and trait anxiety levels during various periods of the experiment. The questionnaires were "author-administered" in order to provide clarification, if needed, to any written instructions and insure that all respondents completed the questionnaires correctly, giving appropriate consideration to each item.

Three types of questionnaire items were used; open-ended, multiple choice, and rating scale. The open-ended items permitted the subject to express his/her answer to the question in one's own words. In all cases, these questions required short (one or two words) objective replies. The multiple choice questions allowed each respondent to choose

the appropriate answer from a list of several options. These multiple choice questions include "dichotomous" items, for example, those requiring only a YES or NO response. Finally, rating scale items were used to obtain judgements or attitudes about some object, concept, or system. These questions permitted the assignment of various response alternatives along an unbroken continuum or in ordered categories along the continuum. Both a graphic scale, allowing the respondent to place his/her judgement any place along the line, and a numerical scale, confining the subject's response to a discrete category along the continuum were employed.

1. User Questionnaire #1

User Questionnaire #1 (Appendix A) employs a combination of question items including open-ended, multiple choice, and graphical rating scale items. Questions 1-22 are designed to obtain information pertaining to occupational, personal and physiological characteristics. Questions 23-40 obtain attitudinal, comparison, and evaluation information pertaining to occupational, operational, physiological and psychological characteristics.

2. User Questionnaire #2

User Questionnaire #2 (Appendix B) utilizes a combination of question items including multiple choice and graphical rating scale items. Questions 1-3 obtained

information relative to physiological factors while questions 4-15 were repetitious items from user Questionnaire #1 designed to obtain attitudinal information from the subjects after using speech recognition equipment for four weeks.

3. STAI Questionnaire

The State-Trait Anxiety Inventory (STAI) is comprised of separate self-report scales for measuring two distinct anxiety concepts: state anxiety (A-State) and trait anxiety (A-Trait). This inventory was developed by Spielberger et. al. at Vanderbilt University and later continued at Florida State University. It was reproduced with the special permission of the Publisher, Consulting Psychologists Press, Inc., Palo Alto, California.

The STAI A-Trait scale consists of 20 statements (Appendix C) that ask people how they generally feel. The A-State scale also consists of 20 statements (Appendix D) but the instructions require subjects to indicate how they feel at a particular moment in time. The STAI was designed to be self-administered and was given individually to each subject. Complete instructions are printed on each test form for both the A-Trait and A-State scales. There were no time limits imposed for completion of the form. Although many of the items have face validity as measures of anxiety, the inventory was referred to as a Self-Evaluation Questionnaire. Each subject responds to every STAI item by

circling the appropriate number to the right of each item statement on the form. Scoring keys are depicted with each scale in Appendices C and D [Ref. 27].

E. EXPERIMENTAL DESIGN

A three-factor mixed design with repeated measures on one factor was employed in this experiment. In consideration of the wide variety of human factors to be examined, the experiment was designed to allow an analysis of three critical factors (occupational experience with microphones, operational training method and experience) affecting recognition accuracy while simultaneously gathering sufficient data to accomplish subsequent analysis on individual characteristics of speaker variability. The two between variables were microphone experience and training method, The third factor, experience (Week#), was the within group variable. A summary of the experimental design appears in Figure 11.

F. PROCEDURE

1. Training

For the T-600, the training procedure consists of entering 10 passes of each utterance into the voice recognizer. A word list of 100 utterances (Appendix E) was provided the subject, each utterance prompted on the CRT,

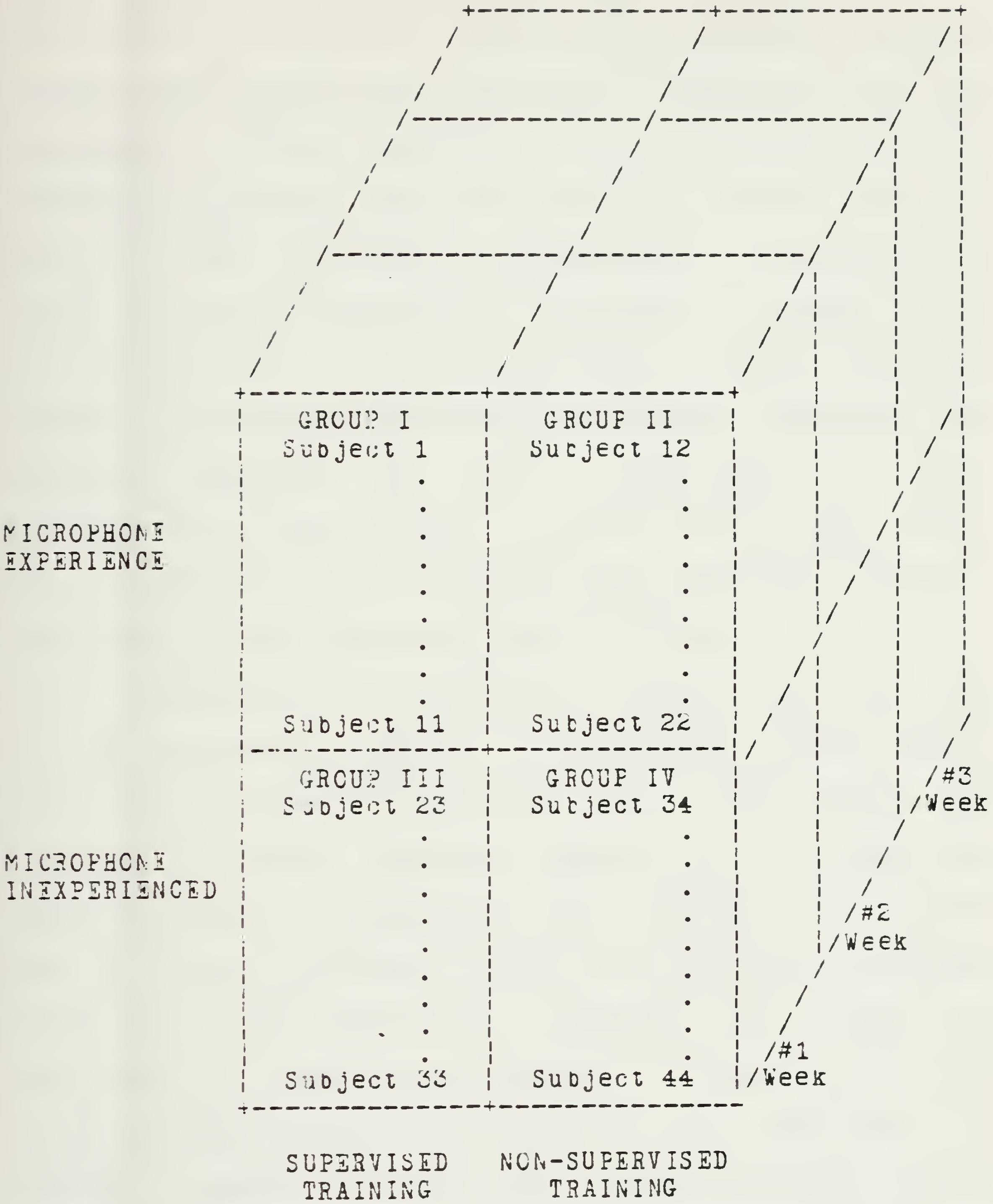

Figure 11. Experimental Design

the 10 passes spoken, and then the next utterance on the list would be prompted. Based on the experimental design, subjects were divided into two groups; supervised and non-supervised. Those supervised during training received detailed instructions, and close scrutiny on each of the 10 passes by the experiment administrator. If the subject failed to clearly pronounce the utterance, if volume level was insufficient, or if the required .1 second pause was omitted, the word was immediately retrained. Non-supervised subjects received the same instructions, a short demonstration of the training procedure and, when ready, were allowed to train the equipment individually with no supervision by the experiment administrator.

Training was accomplished only during the first week of the experiment. Subjects training in the morning (0730-1230 hours) would subsequently test during those periods and likewise for those subjects training in the afternoon (1400-1900 hours). Immediately after training, all subjects made at least two passes of the entire 100 word vocabulary (similar to a test session) to identify any problems in training of a particular utterance. If the utterance was correctly identified on both passes it was considered as trained. However, if an error (either misrecognition or non-recognition) occurred, a third pass was made. If less than two of the three passes of any utterance was correct, that utterance was retrained.

After the equipment was trained, each subject was measured for vital capacity and peak flow rate. Finally, User Questionnaire #1 was administered. Total time for the training session averaged 1.5 hours per subject.

2. Recognition Testing

Following training, subjects were tested on the system. Each subject made 2 passes through the entire vocabulary list on each of three days during the week. Duration of the experiment was three weeks. During Week #1 the vocabulary list remained in the same order as during training (Appendix E) while in Week #2 the order of the utterances were reversed (Appendix F) and in Week #3 the order was randomized (Appendix G). The purpose of this change in vocabulary order was to reduce the effect of learning due to repetitiveness, and thereby provide a more realistic picture of speaker variability. Data was collected in the form of recognitions, misrecognitions, and non-recognitions using Appendix H.

The STAI questionnaire for A-State scale measurement was administered just prior to the first testing session (Week #1, Trials 1-2) to determine anxiety levels prior to using voice equipment. During Week #2 another STAI questionnaire for A-State scale was administered following the first test session of that week. The final STAI form for the measurement of A-Trait scales, was administered

during week #3. User Questionnaire #2 was provided to each subject at the conclusion of the experiment.

3. Vocabulary

It was desired that a test vocabulary similar to a vocabulary intended for practical application in a military environment be used. Of concern in the design of the vocabulary was the fact that brief monosyllabic words are more difficult to recognize that longer polysyllabic words or phrases. A relatively equal distribution of words and utterances containing a syllabic content ranging from 1 to >5 syllables was selected as the final vocabulary. The words were chosen both from previous experimentation [Ref 23] and the author's military experience. Appendix I provides a listing of the 100 utterances used in the experiment and considered as representative of use in a military command center.

G. VARIABLES

The dependent variables in this experiment were total errors, a linear combination of misrecognitions and non-recognitions. Independent variables in the overall experimental design are experience, job function, and training method. Additional independent variables included each of the individual human factor characteristics elicited earlier.

Data was collected on the eleven subjects within each group of the experimental design. Each subject made 600 utterances per week for a grand total of 1800 for the experiment. Total utterances for the completed experiment numbered 79,200 (44 x 1800).

V. ANALYSIS AND RESULTS

A. GENERAL

All analyses were performed using the MINITAB statistical package [Ref. 28]. Repeated measures analyses of variance procedures were performed in accordance with guidance provided by Bruning and Kintz [Ref. 29]. Nonparametric tests for significance between pairs of means, several independent samples, and for trend analysis were conducted utilizing procedures discussed by Conover [Ref. 30]. Additional parametric analysis followed procedures prescribed by Ott [Ref. 31].

All mean error rates that appear in figures are of untransformed data. Since the F test in an analysis of variance is vàlid even with mild departures from the assumption of equality of variances [Ref. 31: p. 630], Hartley´s Test for homogeneity of population variances was used to determine whether an extreme case (unequal variances) existed and thereby determine if a transformation of data would be required to stabilize the variances. Results of this test are presented in Table IV. The assumption of equal variances is the basis for the use of untransformed data in all subsequent analyses.

The correlation coefficient reported herein is Spearman´s Rho. Although the Pearson Product Moment

TABLE IV

TEST FOR EQUALITY OF VARIANCES

DATA:

s^2(group I) = 1947.42
s^2(group II) = 3666.80
s^2(group III) = 2625.82
s^2(group IV) = 5636.95

HYPOTHESES:

H_0: All population variances are equal

H_1: Not all population variances are the same

TEST STATISTIC:

$$F_{Max} = \frac{s^2_{Max}}{s^2_{Min}} = 2.895$$

DECISION:

Level of significance: .05

Tabulated value of F_{Max} = 5.67

CANNOT REJECT THE NULL HYPOTHESIS

correlation coefficient ´r´ is most commonly reported, it is however, a random variable, and as such has a distribution function. Conover [Ref. 30] states that ´r´ has no value as a test statistic in nonparametric tests unless the distribution is known.

B. OCCUPATIONAL CHARACTERISTICS

1. Hypotheses

The following hypotheses pertaining to the occupational characteristics of speakers using voice recognition equipment were tested:

a. H_0: Job function (microphone experienced users versus non-microphone experienced users) will have no affect on recognition accuracy.

H_1 : Job function (microphone experience) affects recognition accuracy.

b. H_0: The branch of service the military member belongs to will have no affect on recognition accuracy.

H_1 : Recognition accuracy is influenced by the branch of service of the user.

c. H_0: A user's attitude pertaining to his/her present job satisfaction will have no affect on recognition accuracy.

H_1 : Job satisfaction affects recognition accuracy.

d. H_0: The degree of satisfaction a user derives from being a member of the military will not affect recognition accuracy.

H_1: Service satisfaction has an affect on recognition accuracy.

e. H_0: The amount of previous computer experience a user has had will not affect recognition accuracy.

H_1 : Previous computer experience affects recognition accuracy.

f. H_0: Competency in a foreign language (bi- or multilingual) will have no affect on recognition accuracy.

H_1: Competency in a foreign language will affect recognition accuracy.

2. Job Function

The results of the experiment for users with and without microphone experience are shown graphically in Figure 12. Microphone experienced users fared only slightly better than non-microphone experienced users. The analysis of variance (ANOVA) results in Table V substantiate this showing an F ratio of .377 indicating no statistically significant difference in the user's job function. Thus, the null hypothesis cannot be rejected.

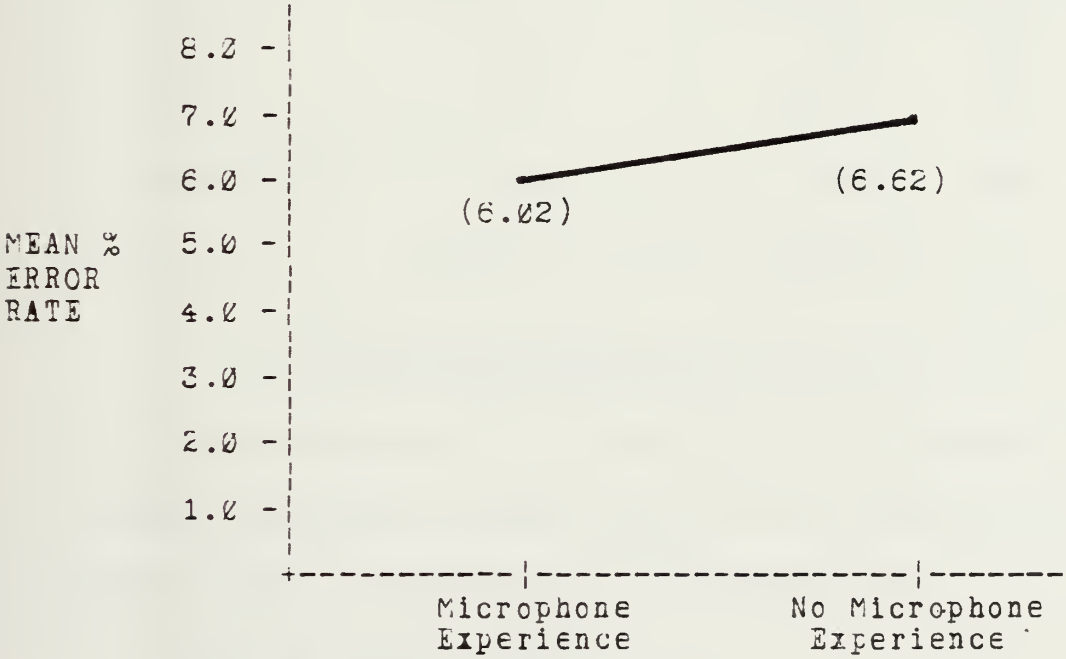

Figure 12. Mean Error Rate vs. Job Function

TABLE V

ANALYSIS OF VARIANCE FOR RECOGNITION ACCURACY

SOURCE	SS	df	MS	F	p
TOTAL	73296.00	131	--	-	--
BETWEEN SUBJECTS	54082.60	43	--	-	--
Microphone Experience (MIC)	436.81	1	436.81	.377	NS
Training Method (TNG)	5629.50	1	5629.50	4.868	**
MIC x TNG	1759.69	1	1759.69	1.521	NS
Error(b)	46256.60	40	1156.41	-	--
WITHIN SUBJECTS	19213.41	88	--	-	--
Trials (TR)	4324.19	2	2162.09	11.696	**
TR x MIC	13.50	2	6.75	.037	NS
TR x TNG	74.32	2	37.16	.201	NS
TR x MIC x TNG	13.00	2	6.50	.035	NS
Error(w)	14788.40	80	184.85	-	--

[** SIGNIFICANT at p < .05]
[NS: NOT SIGNIFICANT for p < 0.05]

Microphone Experience: Experienced vs. Non-experienced

Training Method: Supervised vs. Non-supervised

Trials: Week #1 (Words 1-100)
Week #2 (Words 100-1)
Week #3 (Words in random order)

Mean total error rates for microphone and non-microphone experienced users is summarized in Table VI. The definitive decrease in error rates by time will be discussed later in the review of operational characteristics.

TABLE VI.

MEAN TOTAL ERROR RATES FOR JOB FUNCTION BY WEEKS
(in Percent)

	MICROPHONE EXPERIENCE	NO MICROPHONE EXPERIENCE	$\bar{X}$ WEEKS
WEEK #1	7.04	7.78	7.41
WEEK #2	6.23	6.71	6.47
WEEK #3	4.79	5.39	5.09
$\bar{X}$ JOB FUNCTION	6.02	6.63	6.32

3. Branch of Service

Three branches of service were represented in the experiment with civilian subjects categorized as a fourth branch. A Kruskal-Wallis test for $k > 2$ samples was used to determine if any differences existed. Table VII provides the synopsis of results. The null hypothesis, that branch of service will not affect recognition accuracy, is clearly rejected. Multiple comparisons were made to determine between which pairs of means the differences occurred. The results of this test indicated significant differences between Army/Navy and Army/Air-Force. Differences between

Civilian/Army, Civilian/Air-Force, Civilian/Navy and Navy/Air-Force were not significant.

Further inspection of these results indicated possible confounding due to experience with voice recognition equipment. All Air Force personnel and 3 out of 8 Navy personnel were experienced users. Segregating the experienced and naive users into separate categories and then reconducting the analysis for affect by branch of service showed no statistical significance (Table VII). Using the original hypotheses established, the null cannot be rejected in either the naive only or experienced only cases. Mean error rates by branch of service for all, naive only and experienced only subjects, are presented graphically in Figure 13.

TABLE VII

AFFECT BY BRANCH OF SERVICE

	ALL SUBJECTS	NAIVE	EXPERIENCED
Type of Test	Kruskal-Wallis	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05	.05
Test Statistic	11.90 **	2.79	.23
Critical Level	.0075	.25	.90
** = Significant at stated level of significance			

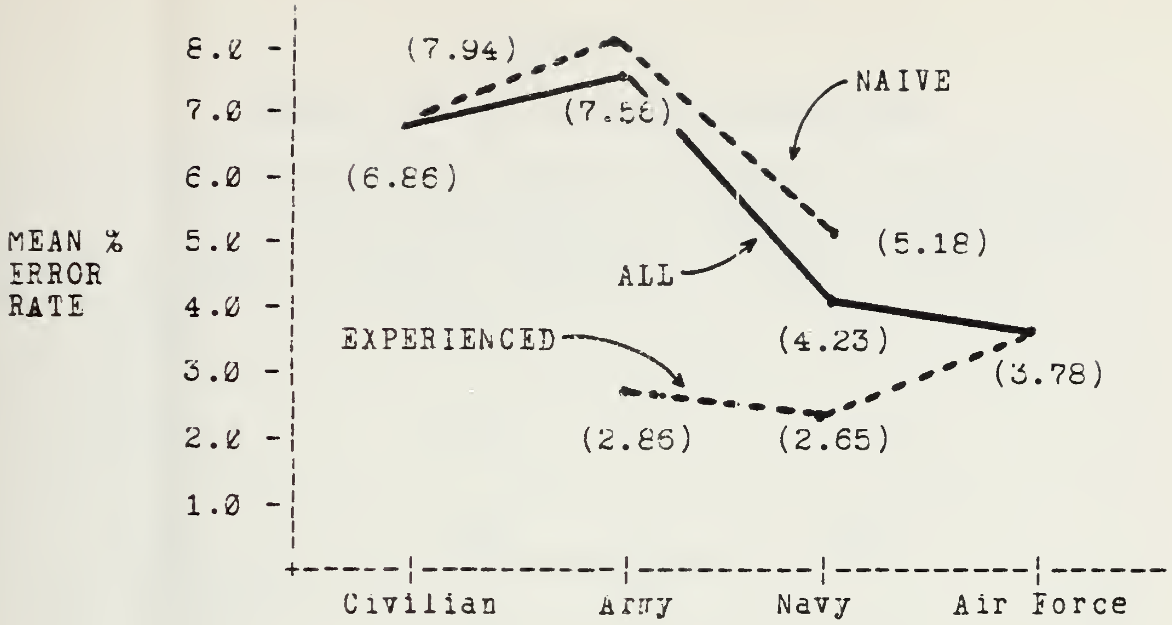

Figure 13. Mean Error Rate vs. Branch of Service

4. Job and Service Satisfaction

Subjects were divided into four groups based upon their subjective responses and included:

a. Persons who disliked their jobs

b. Those who were borderline or neutral in their feelings

c. Individuals who liked their present job

d. Persons who indicated a very definite liking of their job -- liked their job very much

The attained test statistic (Table VIII) leads to the decision that the null hypothesis cannot be rejected. The correlation coefficient between the two variables was not significant and it is concluded that there is no apparent correlation between the satisfaction a user has for his/her

TABLE VIII

AFFECT BY JOB/SERVICE SATISFACTION

	JOB SATISFACTION	SERVICE SATISFACTION
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	4.60	.219
Critical Level	.20	.90
Correlation Coefficient	.016	.041

** = Significant at stated level of significance

current job and how well that user will perform with voice recognition equipment. This particular human factor is nevertheless worthy of further examination in the future in terms of users whose current job entails the day to day use of voice equipment.

In the analysis of the affect service satisfaction has on recognition accuracy, the 2 civilians were removed from the sample population. Subjects were now divided into three groups based upon their subjective responses and included:

a. Those who are unsatisfied or don´t care
b. Those who are reasonably satisfied
c. Those who are very satisfied with their respective service

The test statistic (Table VIII) reveals no significant difference between groups and therefore the null hypothesis, that the degree of satisfaction a speaker derives from being in the armed services will not affect recognition accuracy, cannot be rejected. Correlation between service satisfaction and total error rates, as before, was not significant, thus indicating little or no correlation between the random variables.

5. Previous Computer Experience

Subjects were subjectively divided into four groups based upon their response to question #32 in User Questionnaire #1 and included persons with:

a. No experience

b. Very little experience

c. Some or moderate experience

d. Considerable experience (data processors)

The analysis provided a test statistic (Table IX) which resulted in the rejection of the null hypothesis and the conclusion that previous computer experience will affect recognition accuracy. Multiple comparisons were performed to determine which pairs of means differed. Significant differences occurred between users with, no and considerable experience, very little and moderate experience, and very little and considerable experience. These results demonstrate that possession of experience with data/keyboard input procedures provide a higher recognition accuracy.

Explanation for this occurrence may be attributed to, for example, a data processor's awareness of the time involved for manual entry and the associated error rate as well. The advantages that voice input offers to those computer experienced personnel may well be a psychological or motivational factor in addition to its presence as an occupational characteristic.

These results are further substantiated by the computed correlation coefficient. Performing a one-tail test for negative correlation with the existence of mutual independence as the null hypothesis, we were able to reject this hypothesis and conclude that as computer experience increases, recognition error rates will decrease (Critical Level: << .001). Graphical representation of mean error rates for the four groups are shown in Figure 14.

TABLE IX

AFFECT OF COMPUTER EXPERIENCE

	COMPUTER EXPERIENCE
Type of Test	Kruskal-Wallis
Alpha	0.05
Test Statistic	14.287 **
Critical Level	< .005
Correlation Coefficient	-.516 **
** = Significant at stated level of significance	

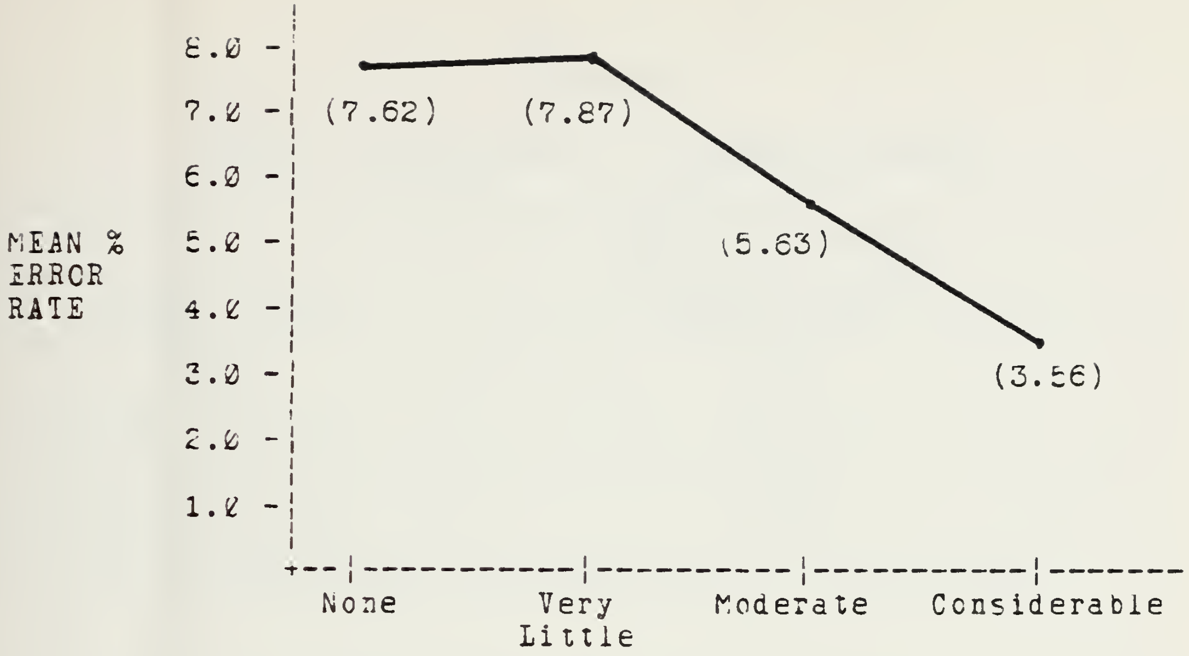

Figure 14. Mean Error Rate vs. Computer Experience

6. Foreign Language Competency

Recognition accuracy was compared between two groups, those with a fluent proficiency in a foreign language and those without. 33 subjects possessed no capability in a second language, whereas 11 were competent in one or more languages. The median total error rate for both groups was 6.28%. A two-sample non-parametric test, the Mann-Whitney, was performed to detect the existence of any differences between the two groups. The computed test statistic (Table X) clearly shows no significance at the .05 level and therefore, the null hypothesis cannot be rejected. The critical regions for this two-tail test included values of the test statistic less than 673 or greater than 814.8.

TABLE X

AFFECT OF COMPETENCY IN ANOTHER LANGUAGE

	FOREIGN LANGUAGE
Type of Test	Mann-Whitney
Alpha	0.05
Test Statistic	754.5
Critical Level	.3776
** = Significant at stated level of significance	

C. OPERATIONAL CHARACTERISTICS

1. Hypotheses

The following hypotheses apply to the operational characteristics under which the subjects were tested.

a. H_0: The method of training a user for voice recognition operation (supervised versus non-supervised) will not affect recognition accuracy.

H_1: Method of training will affect recognition accuracy

b. H_0: The time of day in which a user trains the equipment will not affect recognition accuracy.

H_1: Recognition accuracy of the user will be affected by the time of day in which he/she trains the voice recognizer.

c. H_0: The period of the week in which the user trains the equipment will not affect recognition accuracy.

H_1: The period of the week in which the equipment is trained will affect recognition accuracy.

d. H_0: Experienced users will acquire the same or greater error rates than inexperienced (naive) users.

H_1: Experienced users will have lower error rates than naive users.

H_0: Recognition accuracy will not be affected by weekly experience.

H_1: A user will demonstrate reduced error rates (decreasing trend) as experienced will voice recognition equipment increases.

e. H_0: The operational ease with which voice recognition equipment may be used will have no affect on recognition accuracy.

H_1 : Ease of use will affect recognition accuracy.

2. Method of Training

The results of the experiment for users receiving either supervised or non-supervised training are depicted graphically in Figure 15. Users who received supervision in the training mode fared significantly better than those who did not. The analysis of variance table (ANOVA) in Table V substantiate this claim, providing an F ratio of 4.868 and a critical level of approximately .035. Thus, the null hypothesis is rejected and we may conclude that the method of training does affect recognition accuracy. Mean total

error rates for supervised and non-supervised users are summarized in Table XI.

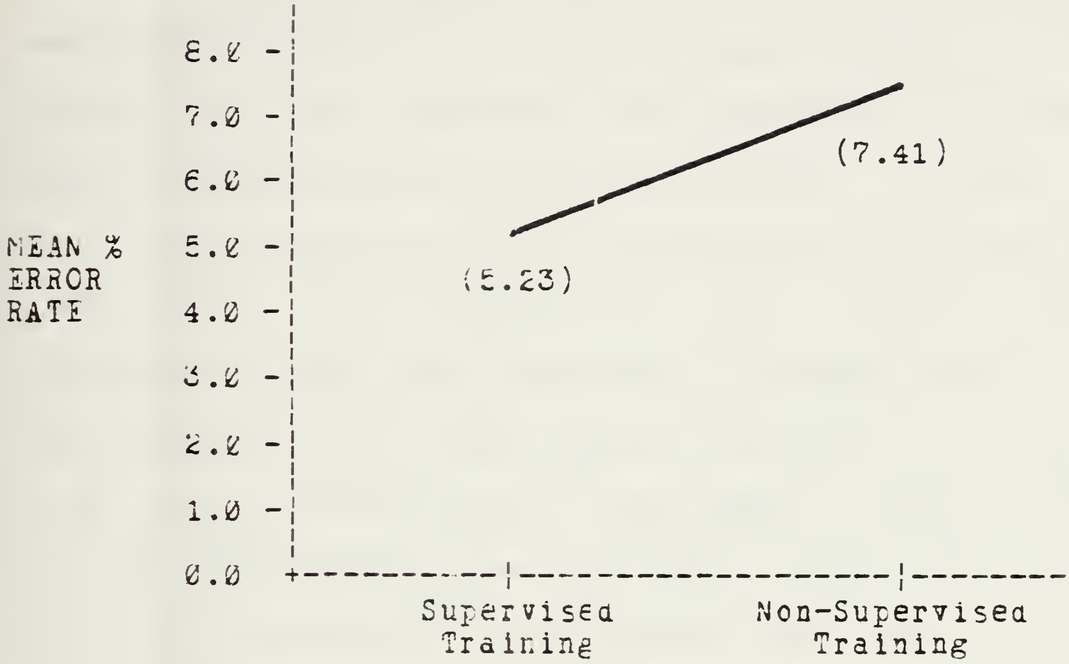

Figure 15. Mean Error Rate vs. Training Method

TABLE XI.

MEAN TOTAL ERROR RATES FOR METHOD OF TRAINING BY WEEKS (in Percent)

	SUPERVISED TRAINING	NON-SUPERVISED TRAINING	$\bar{X}$ WEEKS
WEEK #1	6.21	8.64	7.41
WEEK #2	5.32	7.63	6.47
WEEK #3	4.17	6.00	5.09
$\bar{X}$ JOB FUNCTION	5.23	7.41	6.32

3. Time of Day and Week

Subjects were blocked by time of day; morning and afternoon, and by time of week; early (Monday-Tuesday), mid (Wednesday-Thursday) or late (Friday-Saturday). A Mann-Whitney test was performed to determine if differences existed between the two time of day groups. Morning users had a median error rate of 5.1% while afternoon users had a 6.67% error rate. Because of equal sample sizes, a parametric t-test was performed to confirm results of the non-parametric test. The presented in Table XII will not allow us to reject the null hypothesis. Critical regions for the Mann-Whitney test included values of the test statistic less than 411.5 and greater than 578.5.

With three groups in the time of week variable, the analysis utilized the Kruskal-Wallis test for determination of differences among the groups. The null hypothesis cannot be rejected with a test statistic less than 5.99, for the Chi-square value with two degrees of freedom. The correlation coefficient was found to be significant at the 0.05 level in a test for negative correlation. A premature conclusion that training occurring in the latter portion of the week would yield lower error rates appeared to be counter-intuitive. It was thought that fatigue, and interruption of a weekend would result in poorer training efforts and hence lead to higher error rates in the future. Upon further analysis, this reversed correlation was found

to be the result of possible confounding arising from the large number of experienced users who trained in the later period of the week. Eight out of thirteen late week users were experienced and with their removal from consideration, the correlation between time of week and total error rate became statistically non-significant.

TABLE XII

AFFECT OF TIME OF DAY AND WEEK

	TIME OF DAY		TIME OF WEEK
Type of Test	Mann-Whitney	t-test	Kruskal-Wallis
Alpha	0.05	0.05	0.05
Test Statistic	469	-1.16	4.14
Critical Level	.275	.252	.25
Correlation Coefficient	.093	.093	-2.67 **
** = Significant at stated level of significance			

4. User Experience

Two sets of hypotheses in Section V.C.1.d are incorporated into this phase of the analysis. The analysis of the first set was performed using the Mann-Whitney test and the associated results are summarized in Table XIII. The median error rates for naive users was 7.28% while experienced users attained a 2.75% error rate. Both groups

had equal numbers of supervised and unsupervised users. The correlation coefficient yielded one of the strongest correlations between two variables within the experiment. The null hypothesis can be rejected and it is therefore concluded that experience will affect recognition accuracy.

TABLE XIII

AFFECT DUE TO USER EXPERIENCE

	EXPERIENCE
Type of Test	Mann-Whitney
Alpha	0.05
Test Statistic	869.0 **
Critical Level	< .0001
Correlation Coefficient	-.599 **
** = Significant at stated level of significance	

The analysis of the second hypothesis of V.C.1.d is depicted graphically in Figure 16, (Trials by Job Function) and Figure 17 (Trials by Training Method). In each case no interaction is present, with the weekly error rate showing a steady drop of approximately .8 to 1.4% each week. This graphical interpretation is proven statistically in the ANOVA presented in Table V. That is, the F ratio is well above the 3.11 required for a level of significance of 0.05. The null hypothesis is rejected and it is concluded that

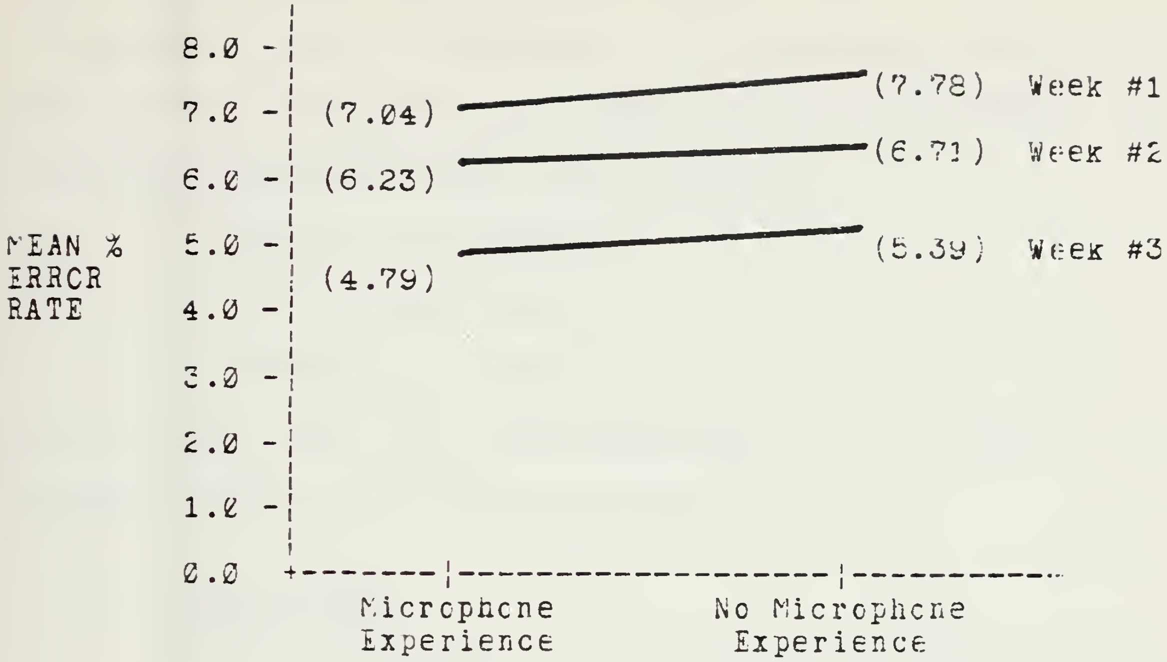

Figure 16. Trials versus Job Function

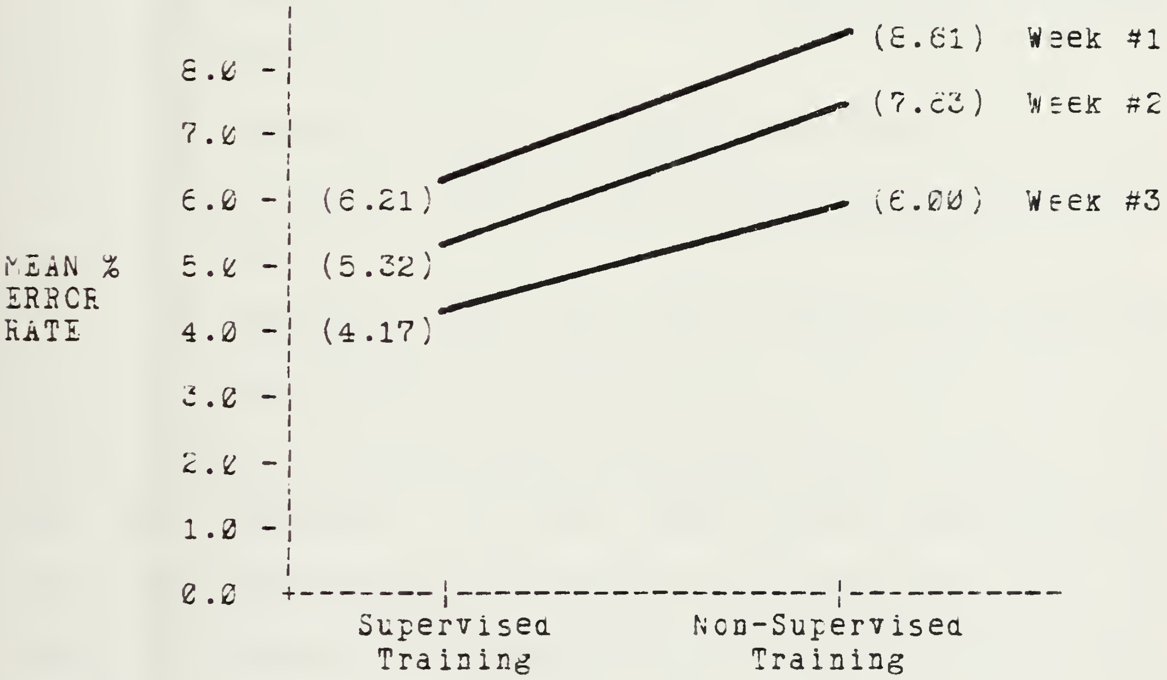

Figure 17. Trials versus Training Method

users will improve (reduce) their error rates through weekly iteration. This conclusion was further verified by application of the Cox and Stuart Test for Trend. The following comparisons were made between:

a. Week #1 and Week #2

b. Week #2 and Week #3

c. Week #1 and Week #3

In all three cases, the null hypothesis, that there is no downward trend, was clearly rejected.

5. Ease of Use

Based on subjective responses by those participating in the experiment four groups were categorized. They include:

a. Users who consider voice recognition equipment difficult to use.

b. Those who had no opinion either way.

c. Users who stated that voice equipment is easy to use.

d. Those who feel that voice recognition equipment is very easy to use.

The results of this analysis are summarized in Table XIV. The test statistic is less than the Chi-square value of 9.488 with three degrees of freedom and therefore the null cannot be rejected. The computed correlation coefficient is not significant at the 0.05 level.

TABLE XIV

AFFECT DUE TO EASE OF USE OF VOICE EQUIPMENT

	EASE OF USE
Type of Test	Kruskal-Wallis
Alpha	0.05
Test Statistic	4.814
Critical Level	> .25
Correlation Coefficient	.157
** = Significant at stated level of significance	

D. PERSONAL CHARACTERISTICS

1. Hypotheses

The following hypotheses were tested pertaining to the personal characteristics of voice recognition users:

a. H_0 : Race of the user will not affect recognition accuracy.

H_1 : A difference in recognition accuracy exists between users of different race.

b. H_0: The marital status of the user will not affect recognition accuracy.

H_1: A user's marital status will have an affect on his/her recognition accuracy.

H_0: Size of a user's family will not affect recognition accuracy.

H_1 : Family size will have an affect on recognition accuracy.

c. H_0: The religious preference/background of a user will have no affect on his/her recognition accuracy.

H_1: A user's religious preference/background will affect recognition accuracy.

d. H_0: A person's accent will not affect his/her recognition accuracy.

H_1: Accent affects recognition accuracy.

e. H_0: The place of birth of a user will have no affect on recognition accuracy.

H_1: One's place of birth affects recognition accuracy.

H_0: The geographic origin of a person will not affect his or her recognition accuracy.

H_1: A person's recognition accuracy will be affected by geographic origin.

f. H_0: The level of education an individual has attained will not affect his/her recognition accuracy.

H_1 : Education level of a user affects recognition accuracy.

g. H_0: The Socio-economic class of a user will not affect recognition accuracy.

H_1: A user's recognition accuracy will be affected by socio-economic class standing.

h. H_0: Past oral-surgery or orthodontal care will not affect recognition accuracy of the user.

H_1: Recognition accuracy of the user will be affected if he or she has undergone oral surgery or orthodontal care.

2. Race

Two racial backgrounds were represented in the sampled population. Thirty-eight Caucasian and six Negro subjects participated in the experimentation. The median total error rate for Caucasian personnel was 6% and 6.8% for Negro users. A Mann-Whitney test was performed to detect the presence of any difference between the two groups. The calculated test statistic (Table XV) was not significant at the .05 level and the null hypothesis cannot be rejected. Critical regions for the test statistic in this two-tail test were values less than 797 and greater than 912.

TABLE XV

AFFECT OF RACE ON RECOGNITION ACCURACY

	RACE
Type of Test	Mann-Whitney
Alpha	0.05
Test Statistic	843.0
Critical Level	.6941
** = Significant at stated level of significance	

3. Marital Status and Family Size

The sample population consisted of 14 single, 25 married, 3 divorced, and 2 other (separated, widowed) personnel. A Kruskal-Wallis test for k > 2 samples was used to determine if any differences in means existed between the

groups. Because the computed test statistic (Table XVI) is less than 7.815, the tabulated chi-square value with 3 degrees of freedom, the null hypothesis cannot be rejected. No correlation coefficient was computed for marital status due to the nominal scale of measurement.

TABLE XVI

AFFECT OF MARITAL STATUS AND FAMILY SIZE

	MARITAL STATUS	FAMILY SIZE
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	2.81	.219
Critical Level	> .3	> .3
Correlation Coefficient	NA	.043
** = Significant at stated level of significance		

The sample population subdivided into five groups for family size with a range from no children to subjects having four or more children. A Kruskal-Wallis test was again used to determine if a difference existed and as before, the null hypothesis cannot be rejected. The computed correlation coefficient indicates mutual independence between family size and total error rate of a voice recognition user.

4. Religious Preference

Although a diverse variety of religious preferences were enumerated by participating subjects, some were pooled to preclude numerous samples sizes of just one person. For example, Methodist and Episcopalian were combined into the Protestant category and so forth. In all, six groups were represented and included Catholic, Protestant, Jewish, Baptist, No Preference and Others (those who could not be readily grouped into one of the aforementioned categories). Using the Kruskal-Wallis test to check for differences between means, the obtained test statistic (Table XVII) does not allow for the rejection of the null hypothesis. Therefore, it may be concluded that the religious preference of the user will not affect his/her recognition accuracy.

TABLE XVII

AFFECT OF RELIGIOUS PREFERENCE

	RELIGIOUS PREFERENCE
Type of Test	Kruskal-Wallis
Alpha	0.05
Test Statistic	3.25
Critical Level	> .25
** = Significant at stated level of significance	

5. Accent

Ten subjects possessed some type of noticeable accent, as determined by the subject and experiment administrator. Seven were Southern and three were categorized as Other (Spanish, Bostonian). Remaining subjects were placed in a 'No Accent' group. The resultant test statistic (Table XVIII) was slightly less than the tabulated Chi-square value of 5.991 with two degrees of freedom. As such, the null hypothesis cannot be rejected. An additional check was accomplished by combining the two accent groups into one generic entity and performing a Mann-Whitney test to detect a difference between the two groups. Again the null hypothesis cannot be rejected at the stated level of significance. Correlation analysis was not performed due to the nominal scale of measurement.

TABLE XVIII

AFFECT OF ACCENT ON RECOGNITION ACCURACY

	ACCENT (3 groups)	ACCENT (2 groups)
Type of Test	Kruskal-Wallis	Mann-Whitney
Alpha	.05	.05
Test Statistic	5.73	704
Critical Level	.055	.09
** = Significant at stated level of significance		

Although the null is not rejected, the critical level is sufficiently close to the stated level of significance. Thus, mean error rates are illustrated in Figure 18 for further examination.

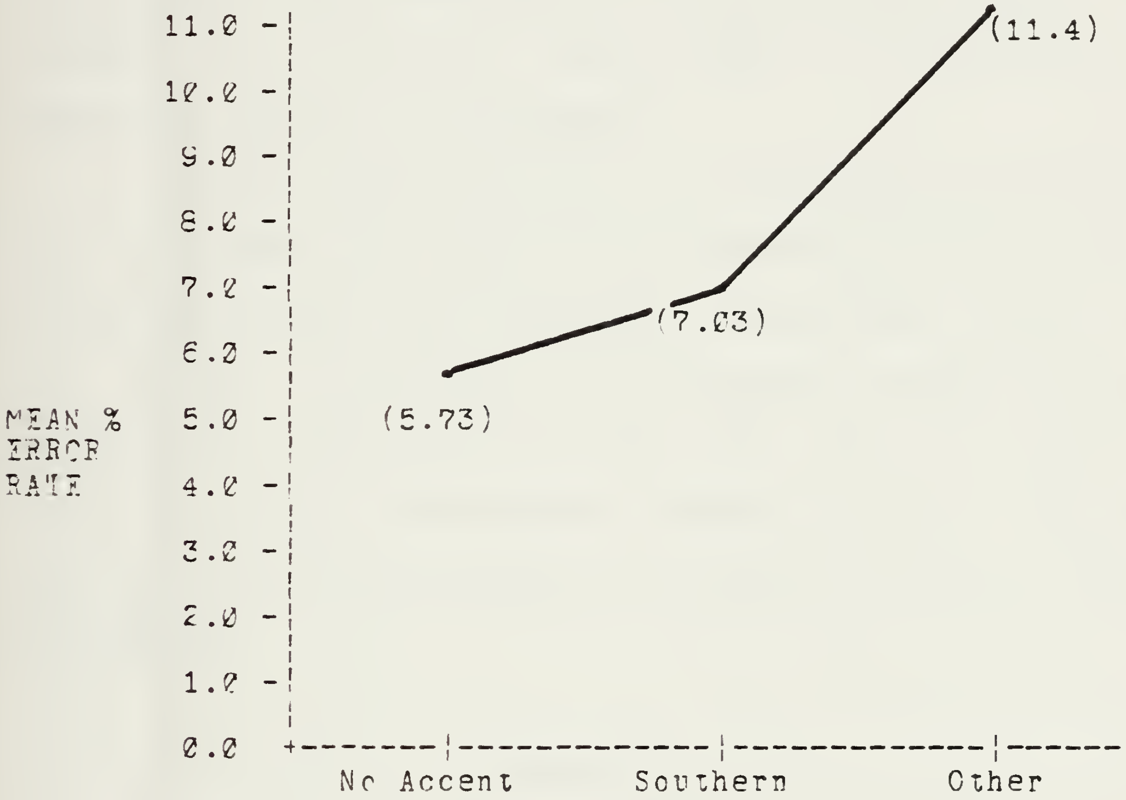

Figure 18. Mean Error Rate vs. Accent

6. Place of Birth and Geographic Origin

Subjects were asked to provide their state of birth and their responses were subsequently classified into one of the following six generic groups:

a. Overseas

b. Northeast United States

c. Southeast United States

d. Mid-Central United States

e. Southwest United States

f. Western United States

Applying the Kruskal-Wallis test to the compiled data, the obtained test statistic (Table XIX) is insufficient to reject the stated null hypothesis.

Because a person's place of birth is not necessarily the environment in which that individual grew up in (ie. during ages 2-18), data pertaining to geographic origin was also tested to determine if any negative affect would be encountered. The geographic areas used were the same as place of birth. Calculated results point to the same conclusion; the null hypothesis of Section V.D.1.e. cannot be rejected.

TABLE XIX

AFFECT OF PLACE OF BIRTH AND GEOGRAPHIC ORIGIN

	PLACE of BIRTH	GEOGRAPHIC ORIGIN
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.25	.25
Test Statistic	5.32	4.09
Critical Level	> .25	> .25
** = Significant at stated level of significance		

7. Level of Education

The sampled population partitioned into the following five categories:

a. High School graduates.

b. Individuals with 1 to 4 years of college but no degree.

c. College graduates.

d. Individuals working toward a graduate degree.

e. Persons accorded a graduate degree such as a Masters or Doctorate.

The data obtained from the five groups was tested for any significant difference between groups. The test statistic (Table XX) leads to the rejection of the null hypothesis and the conclusion that level of education affects the overall error rate for voice recognition users. A relatively strong positive correlation exists with a critical level of 0.006. That is, as the individual increased in level of education, a concomitant decrease in error rate occurred.

Multiple comparisons between the various groups showed the predominant influence to be graduate students. Further examination indicated possible confounding due to that group's prior experience with voice recognition equipment. Eleven out twelve graduate students were

TABLE XX

AFFECT OF LEVEL OF EDUCATION

	EDUCATION (ALL)	EDUCATION (NAIVE)
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	14.300 **	4.18
Critical Level	.015	> .25
Correlation Coefficient	-.380 **	.063
** = Significant at stated level of significance		

experienced users. These experienced users were stripped out of the sample and the Kruskal-Wallis test applied to only those that were naive to voice technology. Using the same hypotheses, the obtained test statistic does not allow for the rejection of the null. This, and the recomputed correlation coefficient corroborate the theory of confounding and the earlier conclusion is now amended to state that level of education will not affect recognition accuracy. Mean error rates for all education levels are shown graphically in Figure 19. Error rates for both, total sample population and naive users only, are included.

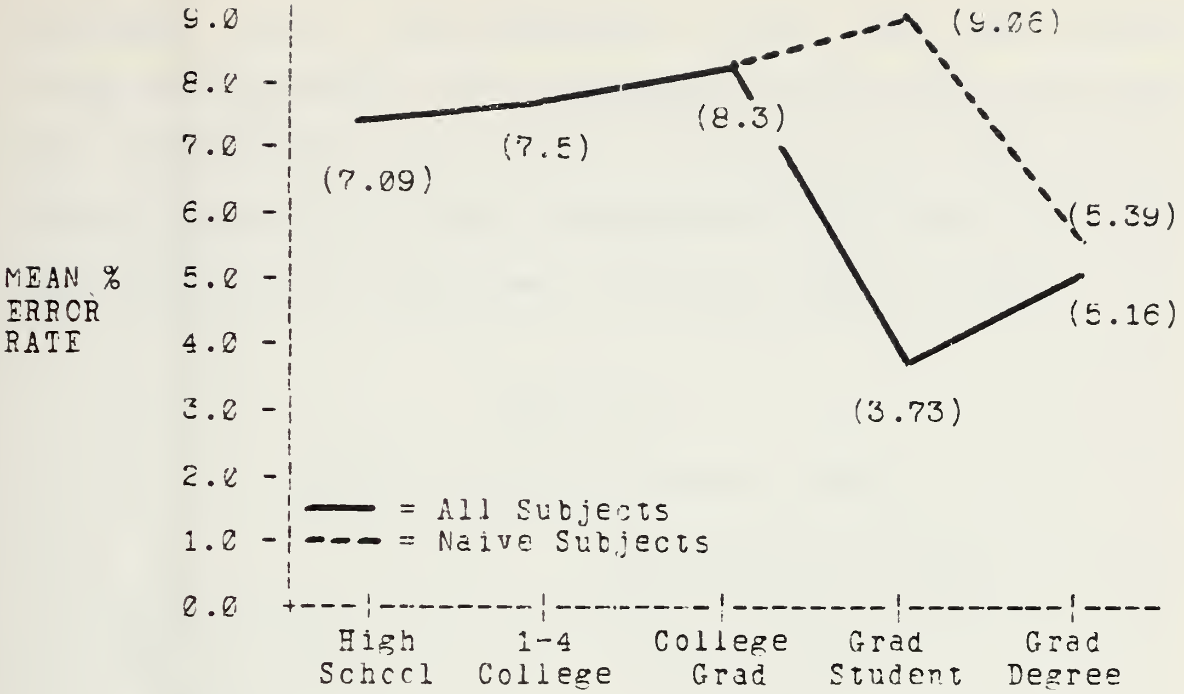

Figure 19. Mean Error Rate vs. Education

8. Socio-economic Class

A variety of socio-economic classes were presented to the participants for selection with one of the following five chosen by each subject:

a. Upper lower class

b. Lower middle class

c. Middle class

d. Upper middle class

e. Lower upper class

The analysis of total error rates for these five groups (Table XXI) yielded a test statistic that would not allow for the rejection of the null hypothesis, and it may be

concluded that socio-economic class will not affect recognition accuracy, The negative correlation indicates that individuals of a lower socio-economic class tend to acquire higher error rates although the coefficient is not significant at the 0.05 level (critical level: 0.158).

TABLE XXI

AFFECT OF SOCIO-ECONOMIC CLASS

	SOCIO-ECONOMIC CLASS
Type of Test	Kruskal-Wallis
Alpha	0.05
Test Statistic	1.95
Critical Level	.83
Correlation Coefficient	-0.152
** = Significant at stated level of significance	

9. Dental

Subjects were queried as to their history of dental care, in particular, oral surgery and/or orthodontal correction. Two groups resulted upon whose data a Mann-Whitney test was performed to determine if any difference existed between them. The null hypothesis cannot be rejected due to the computed test statistic (Table XXII). Critical regions for the test statistic included values greater than 714.69 and less than 635.31.

TABLE XXII

AFFECT OF PAST AND/OR PRESENT DENTAL CARE

	DENTAL CARE
Type of Test	Mann-Whitney
Alpha	0.05
Test Statistic	638.50
Critical Level	.3643
** = Significant at stated level of significance	

E. PHYSIOLOGICAL CHARACTERISTICS

1. Hypotheses

The following hypotheses pertaining to various physiological characteristics of voice recognition equipment users were tested.

a. H_0: The user´s age will not affect his/her recognition accuracy.

H_1: Age will affect the total error rates of users of voice recognition equipment.

b. H_0: The height and weight of an individual using voice technology will not affect overall recognition accuracy.

H_1: Recognition accuracy will be affected by an individual´s weight.

c. H_0: The vital capacity and rate of air flow of a user will not affect his/her recognition accuracy.

H_1: Recognition accuracy will be affected by a person's vital capacity and rate of air flow.

d. H_0: The overall physical condition of the user will not affect his/her recognition accuracy.

H_1: Recognition accuracy will affected by one's physical condition.

H_0: Formal speech and/or voice training will not affect recognition accuracy.

H_1: A user's recognition accuracy will be affected by any formal speech or voice training/therapy.

2. Age

The subjects ranged in age from 20 to 47 and were divided into five groups for purposes of the analysis. These groups and their mean error rates are:

a. 20 to 24 (4.68%)

b. 25 to 26 (7.03%)

c. 27 to 31 (7.15%)

d. 32 to 35 (5.73%)

e. 36+ (6.10%)

These five groups were tested to detect for differences among their means. The obtained results (Table XXIII) show that the null hypothesis, stated above, cannot be rejected and that the two variables, age and total error rate, are mutually independent.

TABLE XXIII

AFFECT ON RECOGNITION ACCURACY DUE TO AGE

	AGE
Type of Test	Kruskal-Wallis
Alpha	0.05
Test Statistic	2.26
Critical Level	> .50
Correlation Coefficient	-0.05
** = Significant at stated level of significance	

3. Height and Weight

Subjects ranged in height from 60 to 77 inches. Four groups were generated for analysis and are listed below with their respective mean error rate.

a. 60 to 64 inches (5.46%)

b. 65 to 69 inches (6.67%)

c. 70 to 72 inches (5.29%)

d. 73 to 77 inches (7.14%)

The results of the analysis, as summarized in Table XXIV, indicate that the null hypothesis cannot be rejected. The small positive correlation coefficient is not significant at the .05 level and thus the variables in question may be considered to be independent.

Weights of the subjects ranged from 110 to 240 pounds. Examination for some natural 'break' points in this range resulted in the creation of the following five groups and their corresponding mean error rates.

a. 110 to 125 pounds (6.48%)
b. 126 to 145 pounds (6.85%)
c. 146 to 175 pounds (5.13%)
d. 176 to 199 pounds (7.18%)
e. 200+ pounds (5.88%)

The null hypothesis cannot be rejected, with the correlation coefficient indicating independence between the two variables.

TABLE XXIV

AFFECT OF HEIGHT AND WEIGHT ON RECOGNITION ACCURACY

	HEIGHT	WEIGHT
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	1.98	1.95
Critical Level	> .50	.75
Correlation Coefficient	.121	.064
** = Significant at stated level of significance		

The similarity in test statistics and correlation coefficients of height and weight may be explained by observing the correlation between height and weight itself. A Pearson product moment correlation of .821 suggests a strong positive association between the two variables and thus serves to confirm the similar results of the analysis.

4. Vital Capacity and Rate of Air Flow

The vital capacity of participating subjects ranged from 1917 to 5725 cubic centimeters. The following four groups were created:

a. 1917 to 2850 cubic centimeters
b. 2851 to 3767 cubic centimeters
c. 3925 to 4450 cubic centimeters
d. 4658 to 5725 cubic centimeters

Analysis for differences between the means of the various groups generated the test statistic (Table XXV) that resulted in the rejection of the null hypothesis. A correlation between increased vital capacity and low error rates was found to be significant using a one-tail test for negative correlation (critical level: .045).

The rate of airflow characteristic had a range of 212 to 731 liters per minute. This range was divided by four and the following groups were used for the analysis. The four included:

a. 212 to 331 liters/min

b. 332 to 460 liters/min

c. 461 to 599 liters/min

d. 600+ liters/min

TABLE XXV

AFFECT OF VITAL CAPACITY AND RATE OF AIR FLOW

	VITAL CAPACITY	RATE OF AIR FLOW
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	8.58 **	6.38
Critical Level	.0375	.095
Correlation Coefficient	-.267 **	-.318 **
** = Significant at stated level of significance		

The test statistic does not allow for the rejection of the null, but a statistically significant correlation coefficient provides an indication that as rate of air flow increases, error rates will decrease. Figures 20 and 21 depict mean error rates for affects due to vital capacity and rate of airflow. Figures 22 and 23 provide the scatter plots upon which the correlation coefficients were determined.

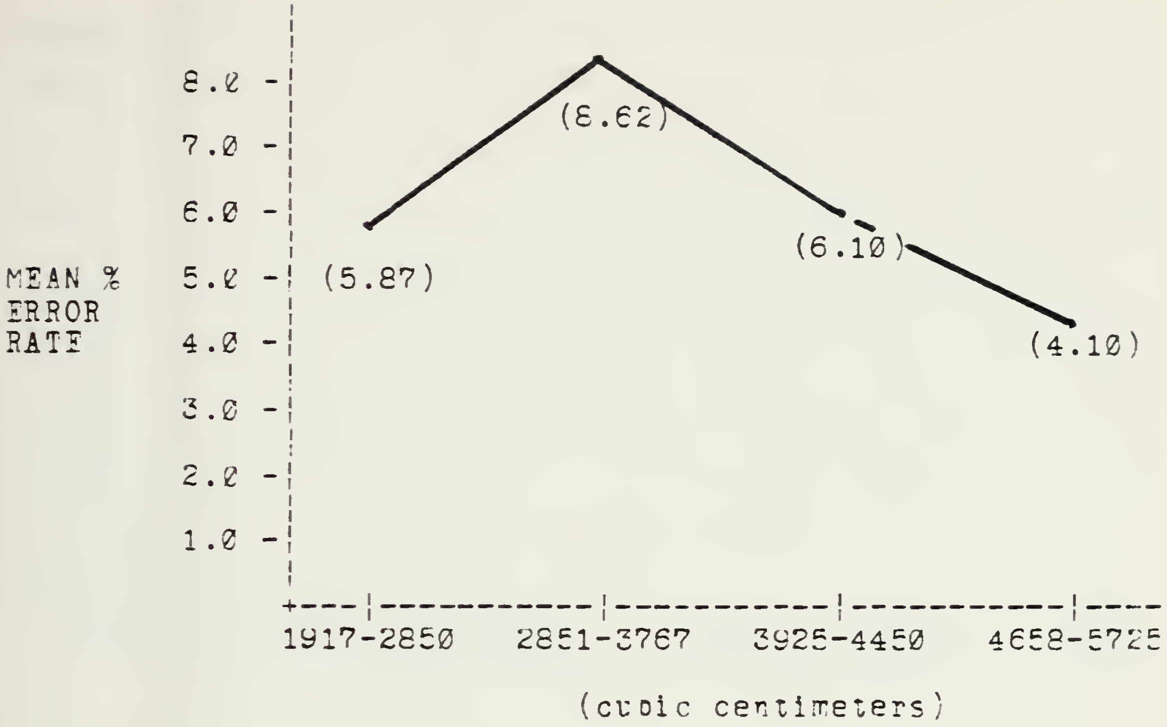

Figure 20. Mean Error Rate vs. Vital Capacity

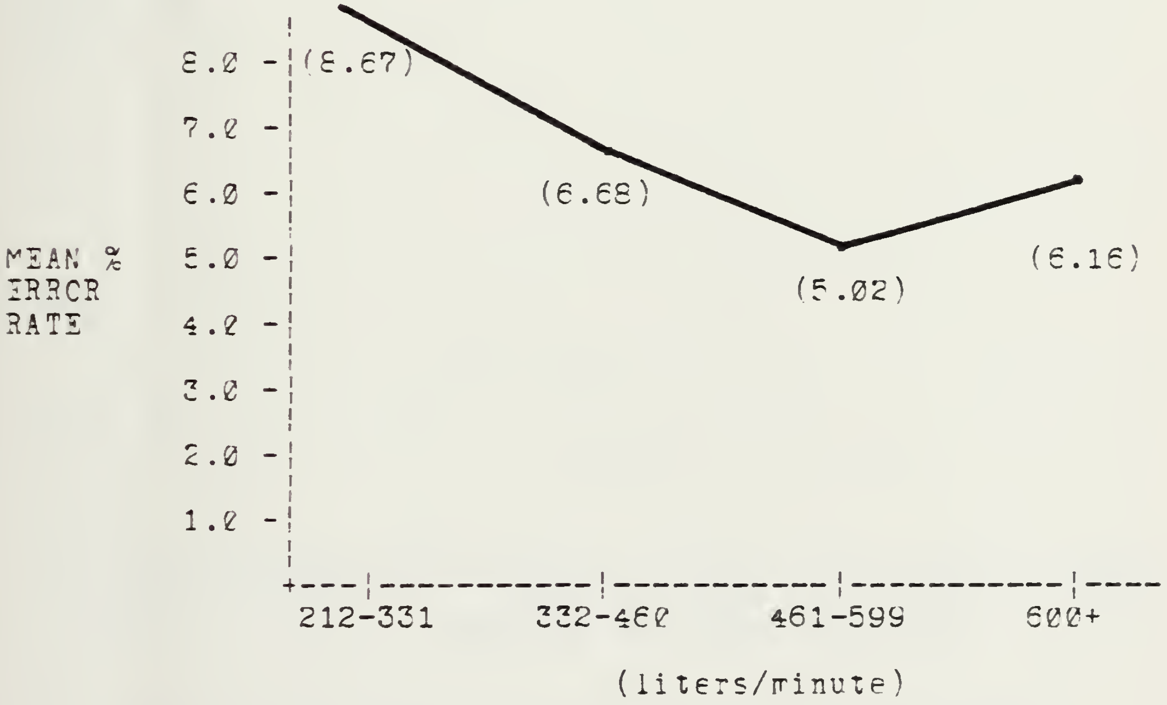

Figure 21. Mean Error Rate vs. Rate of Air Flow

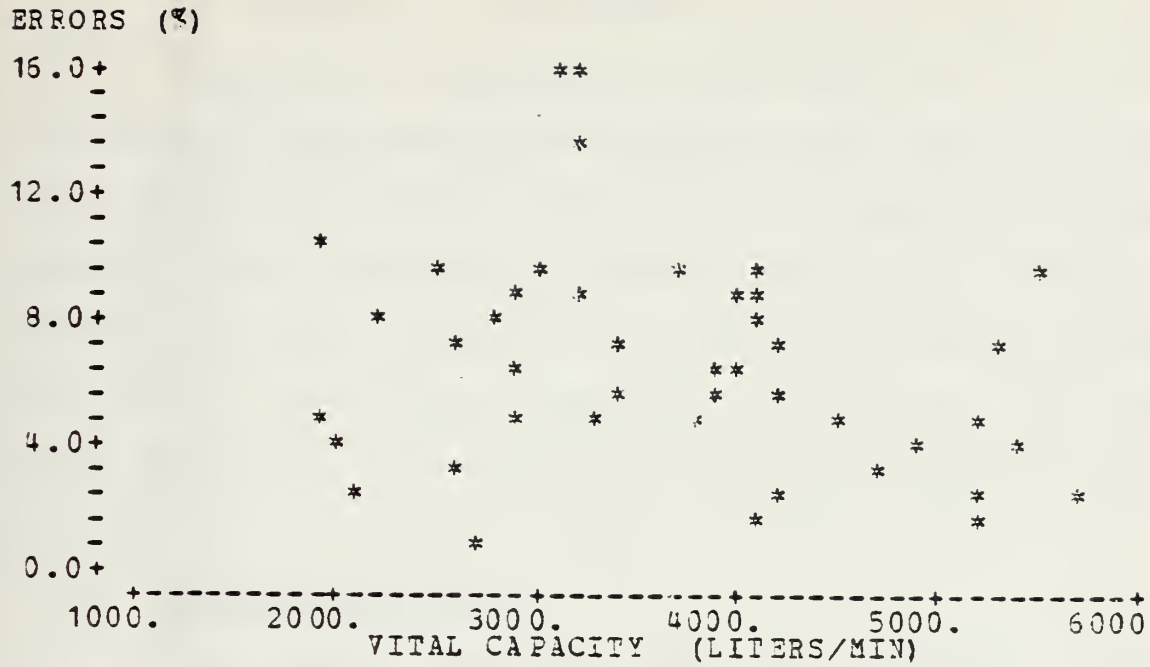

Figure 22. Scatter Plot for Vital Capacity

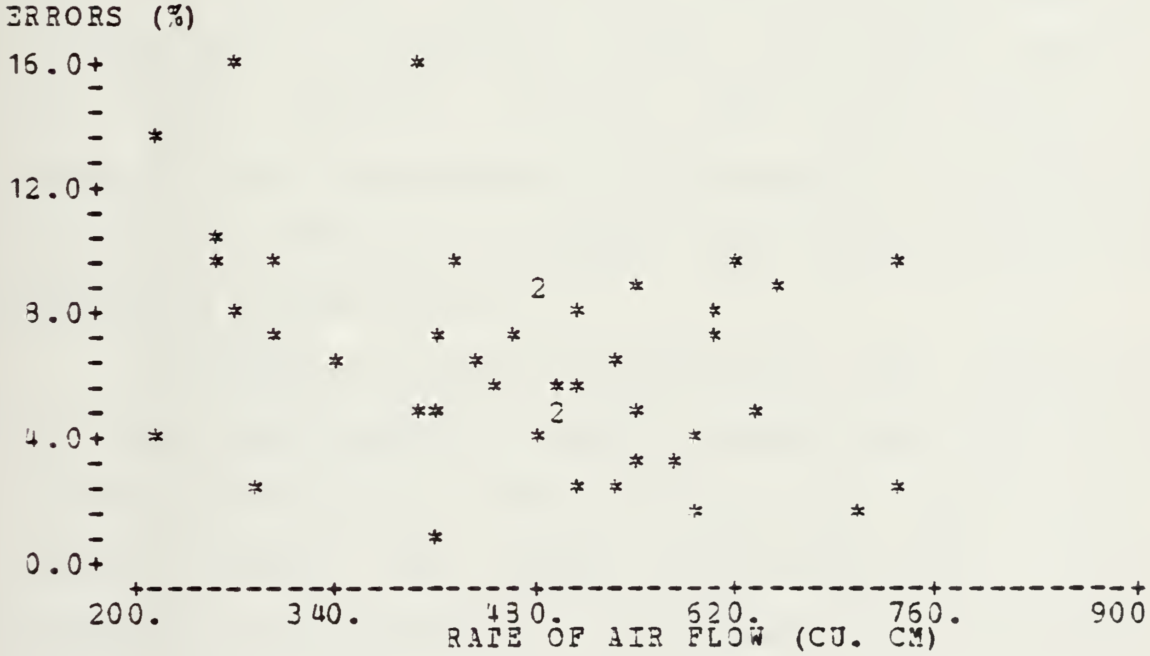

Figure 23. Scatter Plot for Rate of Air Flow

The dilemma of a non-significant Kruskal-Wallis test and a significant correlation coefficient can only be explained by the subjective division of the range of flow rates into the groups used for the analysis. Biased grouping could provide a matrix that would yield a significant test statistic to show a difference between means but in the final analysis, credibility for this characteristic as a determinant in personnel selection would be lost.

5. Physical Condition

Four groups resulted from the subjects' self-appraisal of their general physical condition and include categories of fair/poor, average, good and outstanding physical condition. Their total error rates were examined to determine if a difference between the groups existed. The results presented in Table XXVI do not allow us to reject the null hypothesis. Additionally, a negligible correlation coefficient presumes the two variables to be independent of one another.

Although a subjective response was the determinant for this characteristic, seven subjects who had colds, trained the recognizer. Their condition was such, that a distinct nasality was present while they spoke. A Mann-Whitney test was performed to determine if a difference between the healthy and 'cold' groups existed. The test statistic of Table XXVI further verifies our previous

conclusion; the null cannot be rejected. The critical regions for the Mann-Whitney test correspond to values greater than 893.6 and less than 771.4

Finally, the analysis for affect due to formal speech therapy or voice training resulted in a test statistic that would not allow for the rejection of the null hypothesis, that speech therapy or voice training will not affect a user's recognition accuracy. Critical regions corresponded to values greater than 835 and less than 695.

TABLE XXVI

AFFECT ON RECOGNITION ACCURACY DUE TO PHYSICAL CONDITION

	PHYSICAL CONDITION	SPEECH TRAINING	COLD
Type of Test	Kruskal-Wallis	Mann-Whitney	Mann Whitney
Alpha	0.05	.05	.05
Test Statistic	2.57	761.00	821.5
Critical Level	.45	.46	.368
Correlation Coefficient	0.03	NA	NA
** = Significant at stated level of significance			

F. PSYCHOLOGICAL CHARACTERISTICS

1. Hypotheses

a. H_0: Anxiety will not affect the recognition accuracy of a user.

H_1: Anxiety will affect the total error rate of a user.

b. H_0: The cooperativeness of a speaker will not affect his/her total error rate.

H_1: Speaker cooperativeness will affect recognition accuracy.

c. H_0: The occurrence of recognition errors will not affect overall recognition accuracy.

H_1: A speaker's overall error rate will be affected by the psychological influence of mis- and non-recognitions.

d. H_0: A speaker's beliefs in voice technology as a time saving job aid will not affect recognition accuracy.

H_1: The attitude a person possesses toward the influence of voice on a computer operator's job and their willingness to use voice because of this influence will affect recognition accuracy.

e. H_0: The attitude a speaker has about computers and information processing will have no psychological affect on recognition accuracy.

H_1: A speaker's psychological attitude concerning automation and data processing will affect recognition accuracy.

2. Psychological Anxiety

The results of the State-Trait Anxiety Inventory are depicted graphically in Figures 24 to 26. Figures 24 and 25

show some indication that individuals with a lower state anxiety acquired fewer errors. The relationship between error rate and trait anxiety, shown in Figure 26, depicts a more randomized occurrence of error rates. Correlation analysis substantiates this in that state anxiety during week #1 is statistically significant with week #2 showing some positive correlation but not significant at the .05 level. There is no significant positive correlation between trait anxiety and error rates.

The obtained STAI scores yielded a normal distribution and equal sample sizes of high and low anxiety users. With the basic assumptions for use of a parametric test met, a two sample t-test was used to detect differences between groups. Additionally, the non-parametric Mann-Whitney test was applied for purposes of further verification, however it does not possess the power of its parametric counterpart. Results of the analysis are included in Table XXVII.

In all cases using non-parametric analysis the null hypothesis cannot be rejected, although the critical level shows the test statistic to be just within the acceptance region. The dichotomy in the trait anxiety analysis is interesting; the more powerful parametric test allows the rejection of the null hypothesis whereas the opposite exists

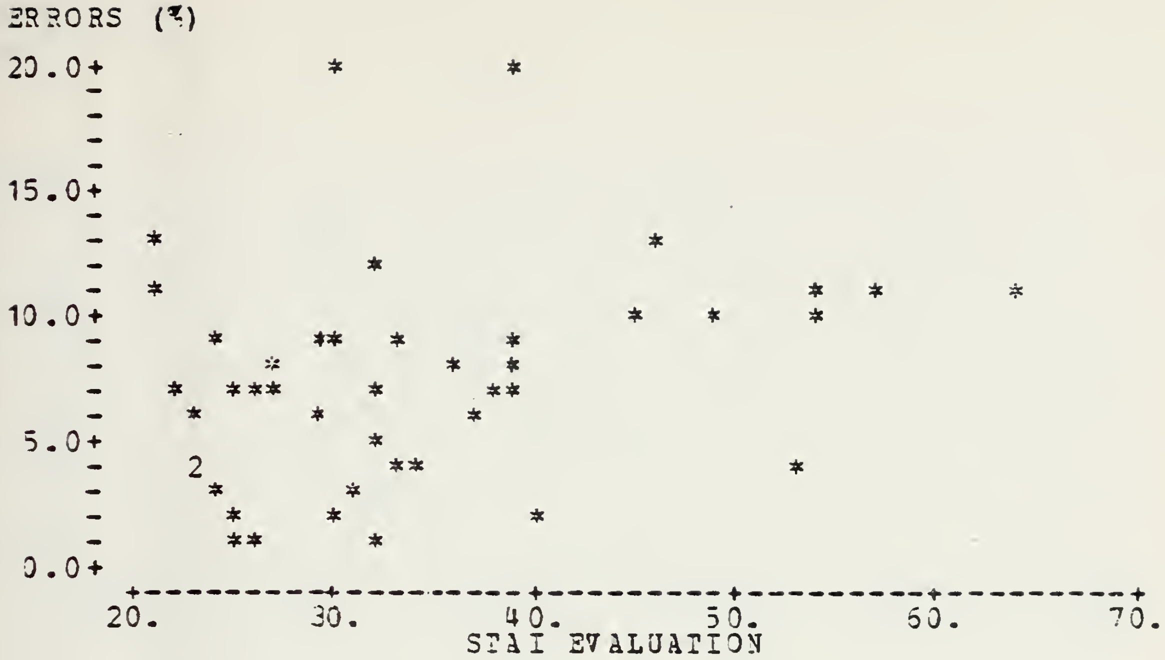

Figure 24. Mean Error Rate vs. State Anxiety (Week #1)

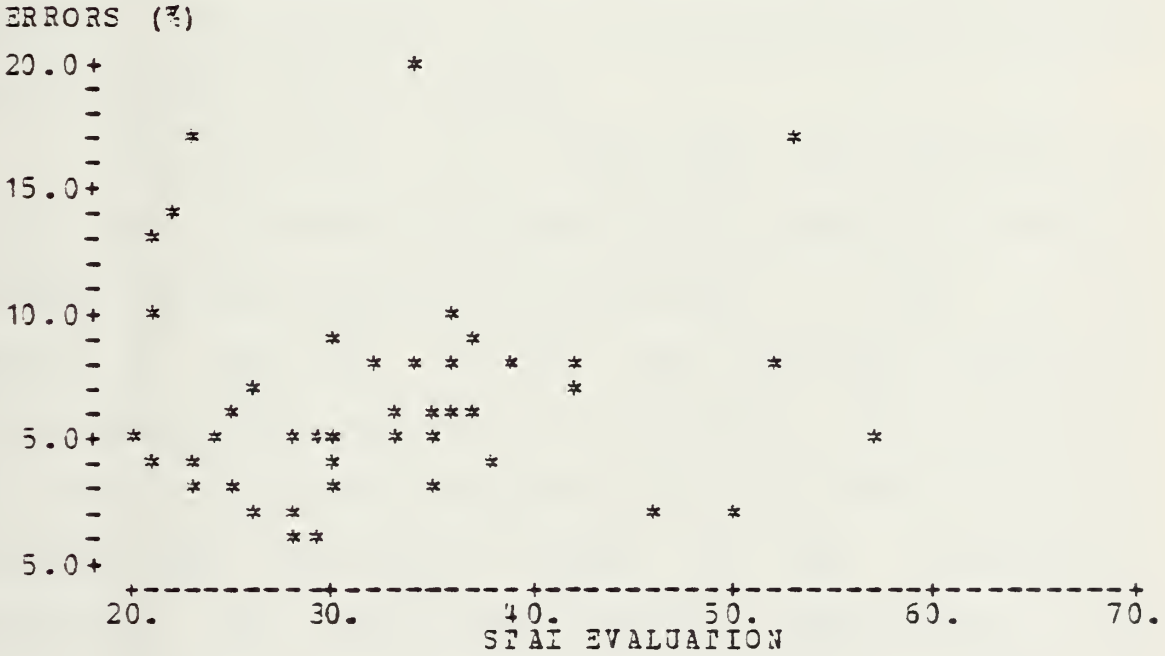

Figure 25. Mean Error Rate vs. State Anxiety (Week #2)

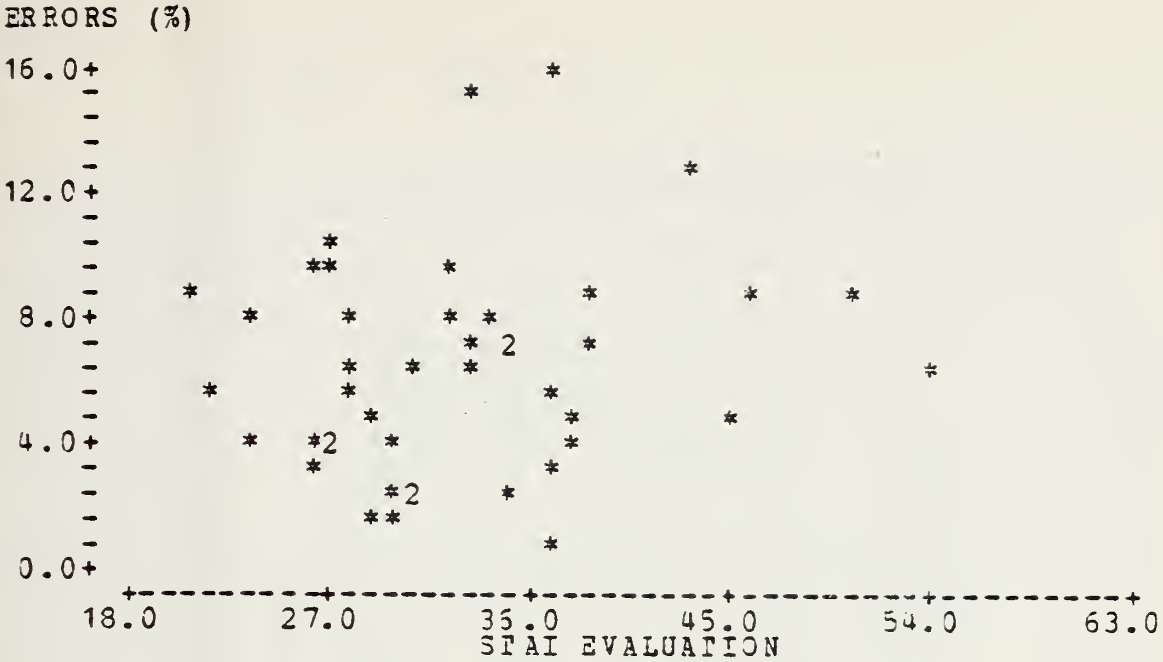

Figure 26. Mean Error Rate vs. Trait Anxiety

using the Mann-Whitney. In both instances though, the test statistic lies extremely close to that point separating the acceptance and critical regions.

The affect due to anxiety may be considered as inconclusive because of the resultant statistical analysis. Although showing significant correlation in Week #1, any anxiety in Week #2 may have been overcome or masked by familiarity and experience with equipment and procedures. By Week #3 and the administration of the Trait inventory, subjects were thoroughly versed in the experimental procedure. The inconsistent results nevertheless, leave reason to believe that anxiety has an affect on speech and hence recognition accuracy, but the degree to which it does remains a clouded issue.

TABLE XXVII

AFFECT ON RECOGNITION ACCURACY DUE TO ANXIETY

	STATE ANXIETY WEEK #1	STATE ANXIETY WEEK #2	TRAIT ANXIETY WEEK #3
Type of Test	t-test Mann-Whitney	t-test Mann-Whitney	t-test Mann-Whitney
Alpha	0.05	0.05	0.05
Test Statistic	-1.313 / 397.5	-1.133 / 420.5	** -2.062 / 419.0
Critical Level	.1966 / .0800	.2639 / .0824	.0461 / .0764
Correlation Coefficient	.326 **	.113	.103

** = Significant at stated level of significance

3. Speaker Cooperativeness

Subjects evaluated their degree of cooperativeness on an interval scale with subsequent creation of the following groups.

a. Less than cooperative speakers

b. Moderately cooperative speakers

c. Very cooperative speakers

d. Extremely cooperative speakers (subjects who marked the ′anchor point′ of the scale)

The results of the analysis are presented in Table XXVIII. with mean error rates graphically represented in Figure 27. The null hypothesis is rejected due to a test statistic greater than the Chi-square value of 7.815. Multiple comparisons among the groups reflect an existent difference between the ′less than cooperative′ and ′extremely cooperative′ speakers only. Despite indication of some correlation between high cooperativeness and low error rate, the computed coefficient is not significant at a .05 level (Critical Level: 0.095).

These results led to a further analysis from a perspective of speaker participation. That is, did the subject like participating in this type of experimentation and if so, could it be correlated to total error rate? Their subjective responses resulted in the creation of three generic groups as follows:

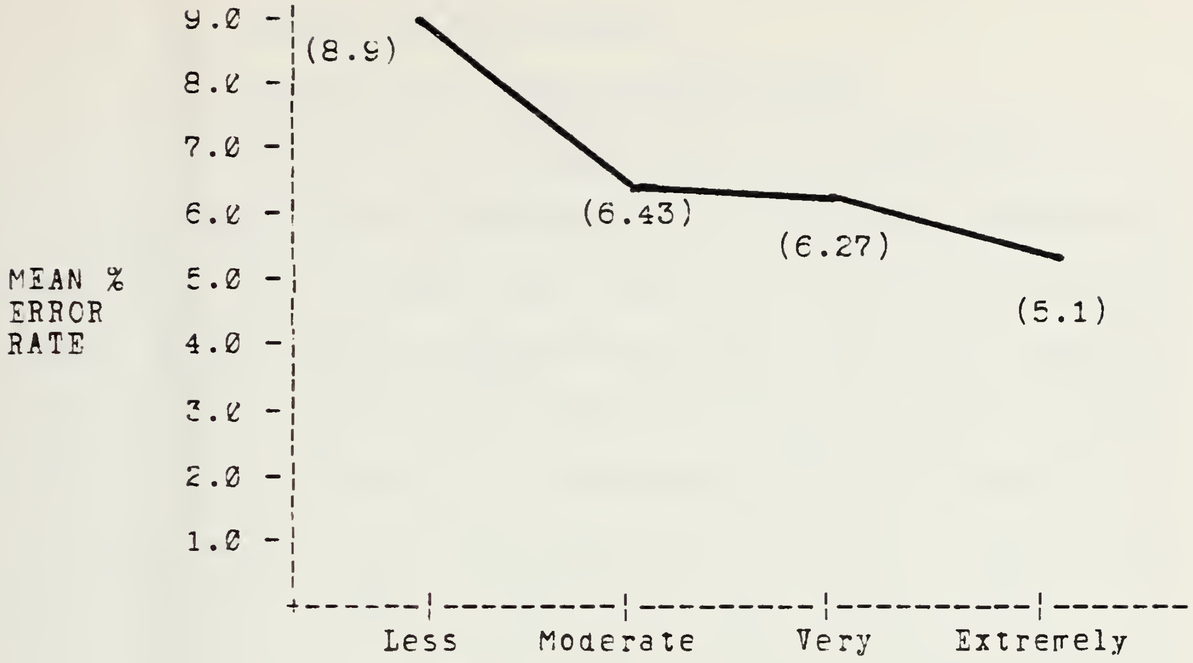

Figure 27. Mean Error Rate vs. Speaker Cooperativeness

TABLE XXVIII

AFFECT OF SPEAKER COOPERATION AND PARTICIPATION

	COOPERATIVENESS	PARTICIPATION
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	16.82 **	4.76
Critical Level	< .005	.095
Correlation Coefficient	-.226	+.278 **
** = Significant at stated level of significance		

a. Those who don´t care

b. Persons who like to participate

c. Persons who strongly like to participate

In this instance the attainment of a positive correlation indicating that those who liked to participate acquire higher error rates is counter-intuitive. The null cannot be rejected based on the computed test statistic given in Table XXVIII. A correlation of .636 between subject responses to cooperativeness and participation is not as large as was expected and as such could, in part, have led to the divergent results. Whether these results are due to willing participants trying too hard to perform well and thus, having greater than usual mis- or non-recognitions is unclear.

4. Recognition Errors

Subjects responded to two questions, one pertaining to their feelings at the time of a mis-recognition and the other pertaining to their feelings over a non-recognition (beep). Their responses to these two questions were averaged to represent how they felt toward the occurrence of an error and this led to the creation of two distinct groups; those who don´t like an error to occur and those who feel they are not disturbed or bothered by an error. The results of the analysis are summarized in Table XXIX.

TABLE XXIX

AFFECT OF RECOGNITION ERRORS

	ERRORS
Type of Test	Mann-Whitney
Alpha	0.05
Test Statistic	612.50
Critical Level	.0897
Correlation Coefficient	-0.225
** = Significant at stated level of significance	

The null hypothesis cannot be rejected and although the negative correlation coefficient indicates that those who dislike errors tend to have higher error rates, it is not significant at an alpha of .05 (Critical Level: .07).

5. Attitudes Toward the Use of Voice

Questions 4, 6 and 8 of User Questionnaire #2 were used to measure the speaker's attitudes toward voice technology. The results (Table XXX) indicate a statistically significant correlation between high error rates and a favorable attitude toward voice recognition as a means of saving time and reducing the burden on a computer operator. Scatter plots of responses to these questions and associated error rates are depicted in Figures 28-30. Multiple comparisons between the groups showed differences between those who would always use voice and those who would

TABLE XXX

AFFECT DUE TO ATTITUDES PERTAINING TO USE OF VOICE

	QUESTION #4	QUESTION #6	QUESTION #8
Type of Test	Kruskal-Wallis	Kruskal-Wallis	Kruskal-Wallis
Alpha	0.05	0.05	0.05
Test Statistic	6.99	9.25 **	7.74 **
Critical level	.075	.025	.02
Correlation Coefficient	.322 **	.437 **	.343 **
** = Significant at stated level of significance			

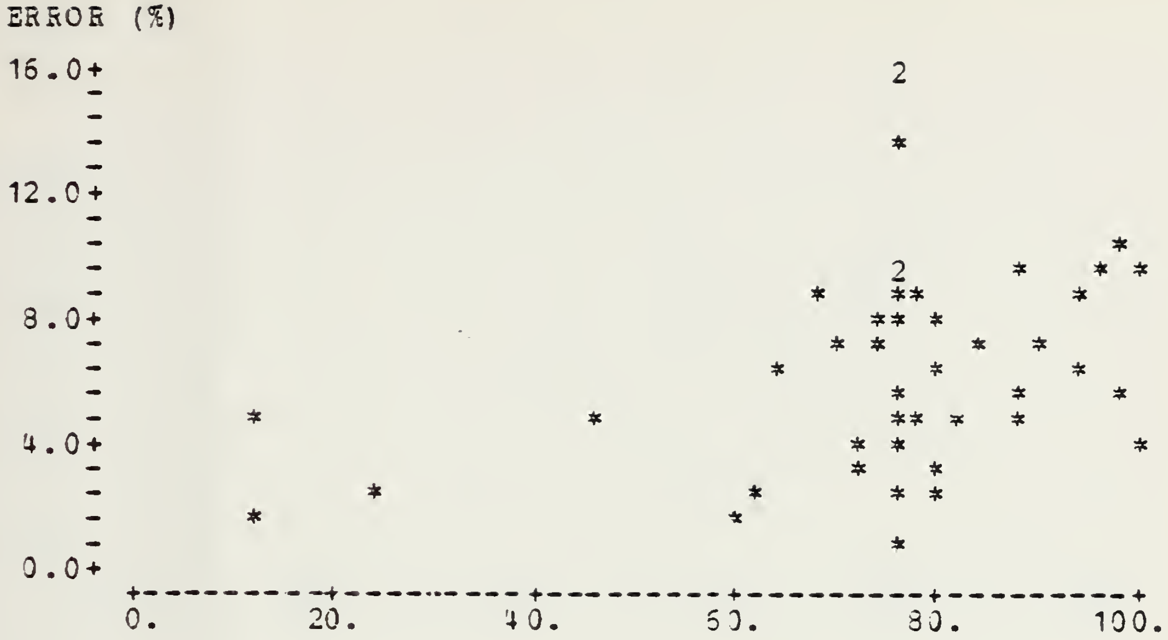

Figure 28. Scatter Plot: Mean Error Rate vs. Question #4

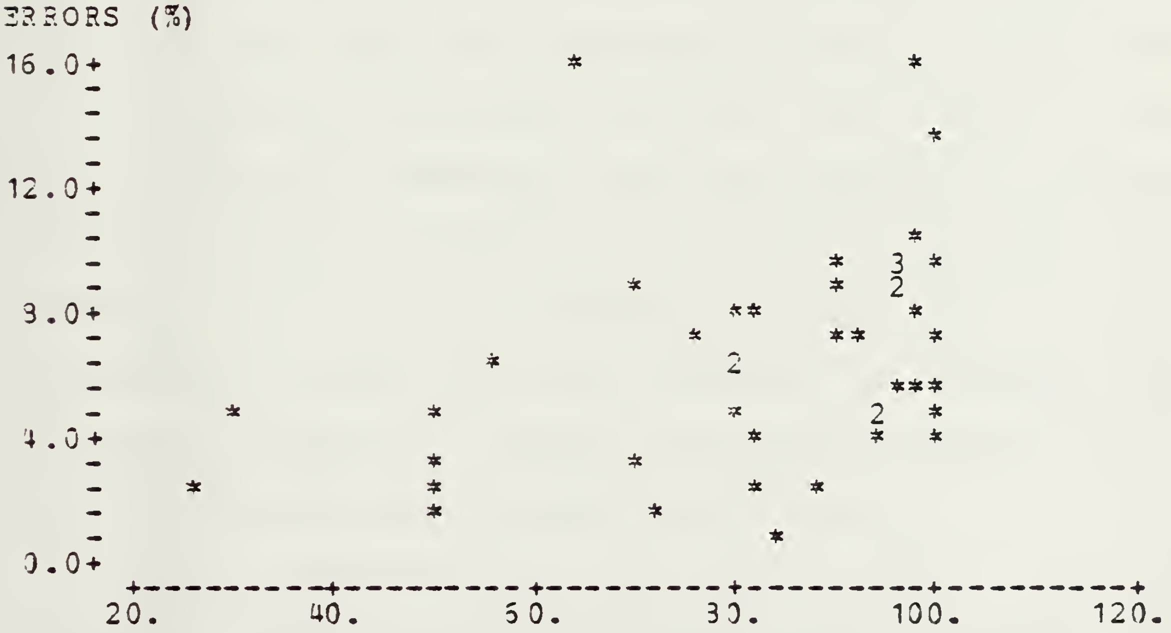

Figure 29. Scatter Plot: Mean Error Rate vs. Question #6

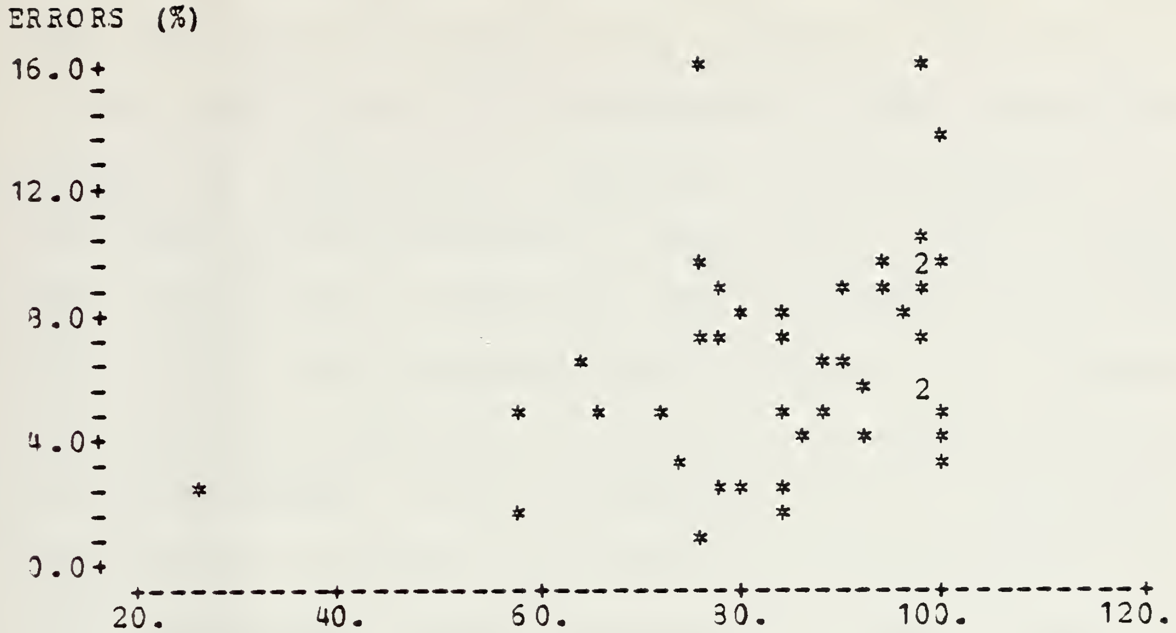

Figure 30. Scatter Plot: Mean Error Rate vs. Question #8

seldom use it despite its pronounced advantages, and between those who felt that the advantages of voice will give the keyboard operator other jobs and those who disagree with such an attitude. Therefore, the null hypothesis cannot be rejected in terms of a speaker's attitude concerning the influence on a data processor's job due to voice recognition. On the other hand, a speaker's willingness to use voice recognition because of his/her beliefs in its requisite advantages will affect error rates.

As was noted earlier, the presence of a positive correlation appears to be contrary to popular belief. One would imagine that a user who believes voice recognition can make the job of a computer operator easier (Question #4),

would tend toward better recognition accuracy. Questions six and eight were asked for the purpose of determining if a user´s error rate might be influenced by the subconscious thought of encountering additional duties because of the efficiency and effectiveness of voice input. But, despite the possibility of additional tasks, potential users still would prefer voice to manual entry. However, the presence of a significant positive correlation may only be attributed to the uniqueness of the situation; ie. as in speaker participation subjects who professed a strong desire to use voice regardless of consequences may have tried too hard for high accuracy and as a result have failed to speak in a ´natural´ manner.

6. Attitude Toward Computers and Information Processing

In response to two sets of questions, subjects provided their attitudes surrounding the necessity of computers in todays society and how voice technology would aid information processing or data input. Attitudes towards computers fell into three general categories.

a. Persons who feel computers are unnecessary.
b. Persons that feel computers are necessary in society, but are not a panacea for all problems.
c. Those who feel that computers are an absolute necessity.

Attitudes toward voice recognition and information processing resulted in four categories.

a. Those believing that voice would take more time for information or data processing.

b. Those with no opinion.

c. Those who feel voice will save some time

d. Those who feel voice can save immeasurable time compared to conventional methods of data entry and information processing.

Results of the analysis are summarized in Table XXXI. Based on these results, the null hypothesis cannot be rejected and thus, it may be concluded that the opinion or attitude a person possesses towards computers, and their feelings pertaining to voice as a time saving advantage will not affect their recognition accuracy.

TABLE XXXI

AFFECT DUE TO ATTITUDES TOWARD COMPUTERS AND DATA PROCESSING

	COMPUTERS	DATA PROCESSING
Type of Test	Kruskal-Wallis	Kruskal-Wallis
Alpha	.05	.05
Test Statistic	.78	3.38
Critical Level	> .8	.15
Correlation Coefficient	.111	-.164
** = Significant at stated level of significance		

G. VOCABULARY ERRORS

As a result of using different numbers of syllables in the vocabulary, it was also possible to get an indication of how well utterances with different numbers of syllables were recognized. Originally done in a longitudinal study [Ref. 24: pp. 9-10] it is analyzed within the context of this document as further verification of those earlier results. This is shown by weeks in Figure 31 and over all conditions in Figure 32. Both figures illustrate a generally declining error rate as a function of the number of syllables in the utterance. Although the current experimentation yielded an approximately 1.5 percent rise in error rate from three to four syllables, it is not a large deviation from the earlier study which indicated little change in error rates between three or four syllables words.

In terms of overall effectiveness, a practical application would dictate the least amount of recognition errors. Therefore, an error rate of 5.91% still remains two to three percent better than utterances with a smaller syllabic content. Despite the higher rate for four syllable compared to five syllable words, the difference is still less than that of one to four or two to four syllables. The variety of vocabulary items used in this experiment further confirms the argument that through a careful and judicious selection of vocabulary items, large vocabulary difficulties and associated high error rates may be reduced.

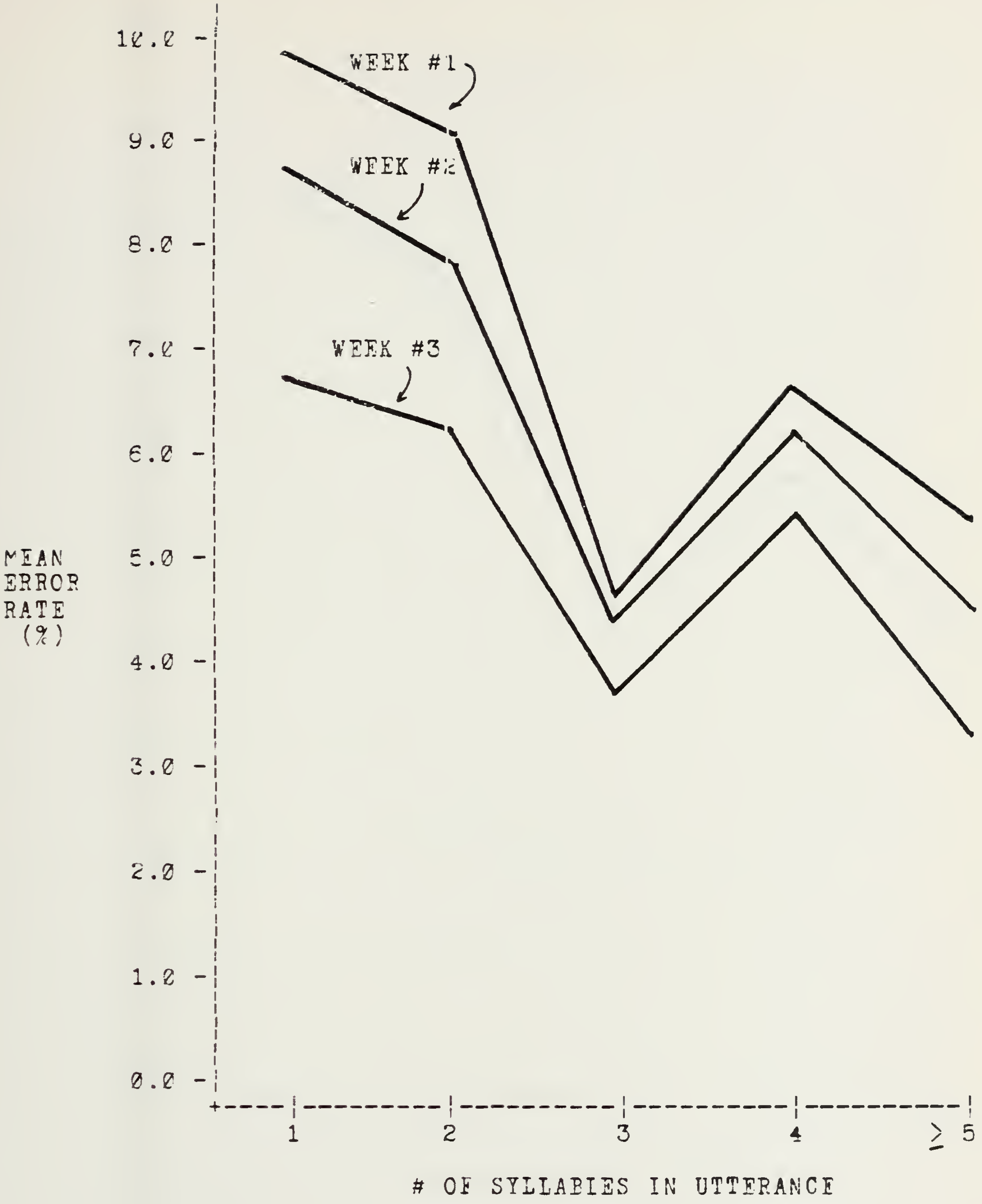

Figure 31. Mean Error Rate vs. # Syllables (by Week)

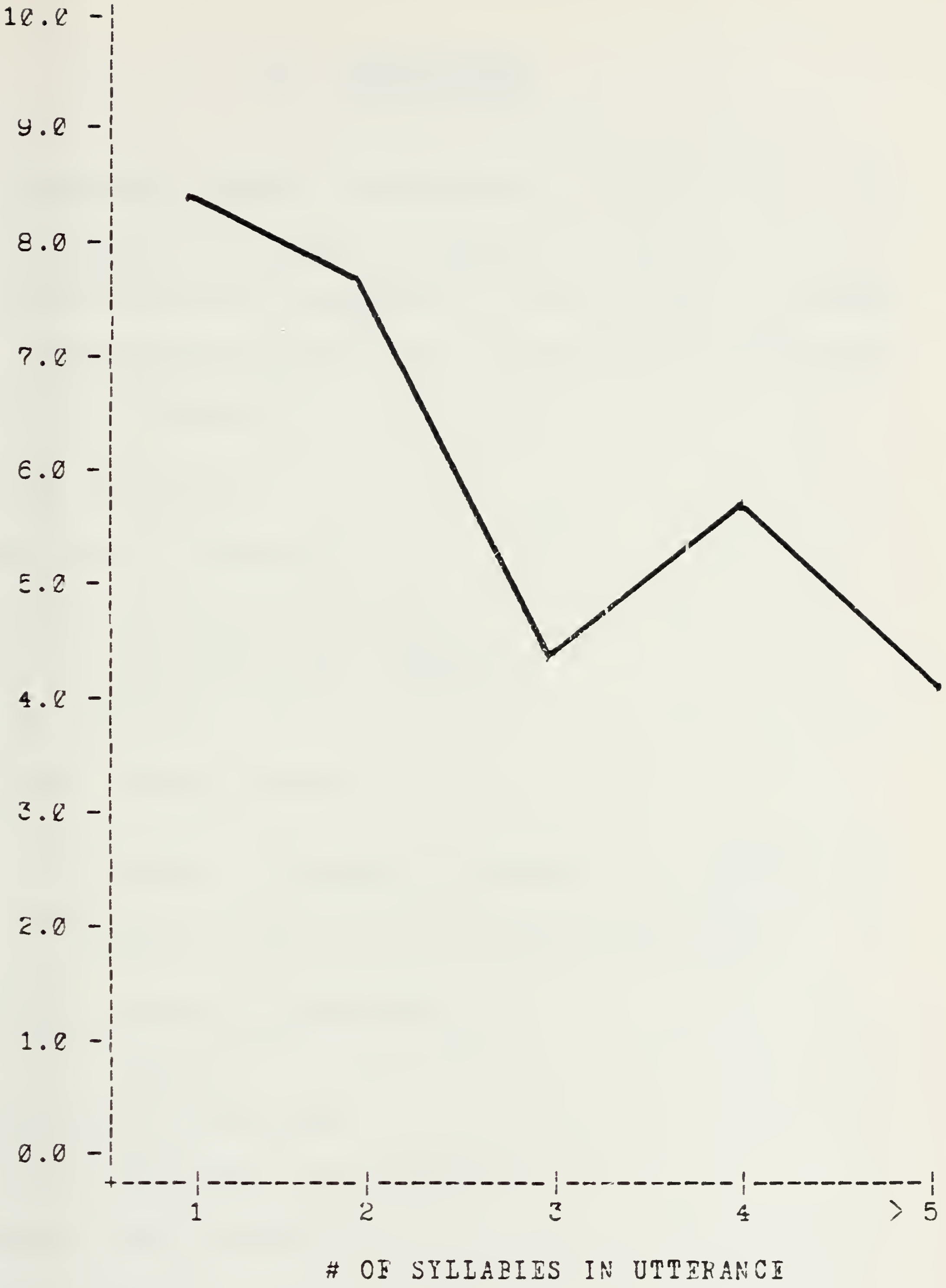

Figure 32. Mean Error Rate vs. # Syllables (Overall)

VI. CONCLUSIONS

Following the lengthy elaboration of results in the previous section it would be helpful to recapitulate, in a brief summary form, the responses of the different variables tested. Variables resulting in a statistically significant test statistic included:

-- Method of training

-- Experience of the user

-- Previous computer experience

-- Level of education (all subjects)

-- Vital capacity

-- Speaker cooperativeness

The following variables produced a significant correlation between itself and recognition error rate.

-- Previous computer experience

-- Time of the week

-- Experience of the user

-- Level of education (all subjects)

-- Speaker participation

-- Vital capacity

-- Rate of air flow

-- State anxiety (first week only)

-- User attitudes pertaining to voice

The following variables resulted in either a non-significant test statistic and/or correlation coefficient.

- -- Job function
- -- Branch of service
- -- Job satisfaction
- -- Service satisfaction
- -- Foreign language competency
- -- Time of day
- -- Time of week (test statistic only)
- -- Ease of use of voice equipment
- -- Level of education (naive users)
- -- Socio-economic class
- -- Dental care
- -- Race
- -- Marital status and family size
- -- Religious preference
- -- Accent
- -- Place of birth/geographic origin
- -- Age
- -- Height and weight
- -- Rate of airflow (test statistic)
- -- Physical conditioning/speech training
- -- Anxiety: State and Trait
- -- Speaker cooperativeness (correlation)
- -- Speaker participation (test statistic)

-- Affect of recognition errors

-- Attitudes toward computers/data processing

The wide range in error rates, .50 to 15.7 percent, for the individual subjects (See Appendix J for a complete summary) indicates an obvious variability between subjects. Within the context of the main experiment and the associated ANOVA, the three variables of job function, training method, and experience (trials), are independent events and are protected from confounding due to the experimental design. The selection of a level of significance equal to .05 is merely to show a possible existence of some effect, not to demonstrate a rigorous test of a stated hypothesis. As the analysis progresses to the extraction of numerous other human factors, these protections and the accompanying power of a parametric test are reduced. In some instances an awareness of a possible dependence between conditions is necessary prior to reaching an ultimate conclusion. For example, were those subsets of a category achieving statistical significance also trained with supervision and/or experienced users and if so, how many were in that particular subset?

The results presented herein suggest that speaker variability would not affect recognition accuracy to such an extent as to preclude its use among only specially selected users. For implementation in military applications, this proves to be especially satisfying since it would negate the

services from the necessity of classifying personnel into particular military occupational specialties or subspecialties for the express purpose of operating voice equipment. It is apparent from the experimentation, and the diversity of skills and experience contained within the sample population, that practically anyone may be a potential candidate to operate voice recognition equipment.

The phrase ´practically anyone´ should be qualified here. Interspeaker variability had a significant impact in the case of one subject, who possessed a severe speech impairment; stuttering. It became obvious in the early stages of training that he would be unable to finish the training phase. In fact, after 30 minutes, only 11 utterances had been satisfactorily placed into memory. Although the individual was eliminated as an experimental subject, his difficulty demonstrates that although most anyone can use this type of technology, there will always exist those, albeit few in number, who for one exception or another are unable to attain a suitable level of recognition accuracy.

The current experimentation has clearly shown that experience and method of training voice equipment can provide excellent recognition accuracy rates. Of course, what determines an ´excellent´ rate is purely subjective and determinate upon the application in which emplaced. What makes this observation readily appealing is that both

characteristics are controlled by the human. They are not factors that one is born with or has inherited. Rather, with closely supervised training procedures, by an experienced operator, a ´naive´ user can quickly attain recognition rates greater than 95 percent and with repetitive experience increase this accuracy until errors are reduced to less than two percent. It must be reiterated that in the present experiment, subjects were not allowed to retrain the recognizer during the three weeks of recognition testing. In actuality, the speaker would retrain an utterance rather than to continue incurring mis- or non-recognition errors.

To a lesser degree, speaker cooperativeness and amount of previous computer experience are definitely factors to be considered. The latter characteristic influences the personnel selection process while speaker cooperativeness, like training and experience, can be influenced by the human element. Certainly, because of data processing experience, such individuals can readily identify with the advantages of speech input and thereby become a more or highly cooperative speaker. Thus combined, these two factors strongly support the potential for achieving high recognition accuracy.

The presence of occasional positive correlation coefficients, that were statistically significant, are difficult to explain or resolve conclusively. Such instances as level of participation, desire to use voice,

and attitudes pertaining to voice, provided misleading results. It was surmised that speakers who are willing participants and find voice to be a technology that they would likely use, would achieve low error rates. The observation to the contrary, supposes that many of those speakers tried too hard for perfect recognition accuracy, and as a result, were less apt to speak naturally. In effect, they were trying to outsmart the machine.

Thus, in an operational environment it becomes incumbent upon both the speaker and the supervisor to fully embrace the concept of voice technology for use in a practical application. In demonstrations at the Naval Postgraduate School it is frequently noted that observers are genuinely impressed with the capabilities of voice input of data until that one error, sometimes after more than 200 successfully recognized utterances, occurs and they sit back and remark that perhaps "additional research is needed prior to placing it into operational use". It is obvious that voice technology is acceptable for use in a military command center and must be fully supported by the Commander and his Staff. If it is, error rates can be minimized by human controls such as training and experience. In conclusion, consistency may best describe the key to speaker variability. Attitudes, training, and experience together, produce consistency in speech and consistency generates a continued high recognition accuracy rate.

APPENDIX A

USER QUESTIONNAIRE #1

NAME:______________________________ SUBJECT#:____________

INSTRUCTIONS:

The purpose of this questionaire is to obtain information from you regarding physical characteristics, personal background, and opinions pertaining to voice recognition equipment and its use. Your answers will assist in determining whether personal and/or physiological traits contribute to effective utilization of voice recognition equipment.

The questions include multiple choice, YES/NO, rating scale and short answer (one or two words ONLY!) types. Appropriate guidance accompanies each question or block of questions.

Your name is NOT required but is requested in order to ease the necessary correlation of your replies with your results in the experimentation. If you desire anonymity, please respond with your subject number only. Please respond truthfully. Check your questionaire after completion to insure you've completed all the questions.

Thank-you for your assistance in this experiment.

In questions 1 - 22, provide either a one or two word response, or place an 'X' by the appropriate answer.

1. What is your age? __________

2. What is your height (in inches)? __________

3. What is your weight? ____________

4. What is your race?

_____ White (Caucasian)

_____ Yellow (Asian/Mongoloid)

_____ Black (Negroid/African)

_____ Red (American Indian)

5. What is your nationality?

_____ Native Citizen of the United States

_____ Naturalized Citizen of the United States

____ Alien

6. What is your religious preference? ____________________
(See Attached Sheet)

7. What is your ethnic background?

_____ Puerto Rican

_____ Filipino

_____ Mexican

_____ Cuban

_____ Latin American (persons from Central or S. America)

_____ Other Hispanic Descent (Extraction not delineated as Mexican, Puerto Rican, Cuban or Latin American)

_____ Eskimo

_____ Aleut

_____ Indian

_____ Melanesian

_____ Chinese

_____ Japanese

_____ Korean

_____ Polynesian

_____ Vietnamese

_____ Other Asian Descent (Extraction not delineated as Chinese, Japanese, Korean, Indian, Filipino, or Vietnamese)

_____ None of the Above

_____ Other (Please specify________________________)

8. Do you have an accent?

_____ YES (what kind? ______________)

_____ NO

9. What is your Marital Status

_____ Married

_____ Divorced

_____ Single

_____ Other (separated, widowed)

10. How many children do you have?

_____ 0

_____ 1

_____ 2

_____ 3

_____ >4

11. Do you wear glasses?

_____ YES

_____ NO

12. Have you ever had orthodontist care &/or wear/worn braces?

_____ YES

_____ NO

13. What is your level of education?

_____ Non High School Graduate

_____ High School Graduate

_____ Associate Degree

_____ 1 year of college

_____ 2 years of college

_____ 3 years of college

_____ 4 years of college (no degree)

_____ College graduate (BA/BS)

_____ Graduate work of more than 1 year (no degree)

_____ Masters Degree received

_____ Doctorate Degree received

14. What state were you born in? ___________________

15. During ages 1-18, in what state did you principally reside? _____________ _________

16. What has been your state of residence for the majority of the last three years? ________________________

17. Do you speak any foreign language(s)?

_____ YES [which one(s) ______________________]

_____ NO

18. What is your branch of service?

_____ Navy

_____ Army

_____ Marine Corps

_____ Air Force

_____ Other (civilian)

19. How many years have you been in the service? ________

20. Have you ever been overseas for more than 13 consecutive months? (not including leave or vacation)

_____ YES (go to question #21)

_____ NO (go to question #22)

21. How many months were you overseas? _______

In what country? ________________________

22. What do you consider to be your socioeconomic class?

_____ Lower Class

_____ Upper Lower Class

_____ Lower Middle Class

_____ Middle Class

_____ Upper Middle Class

_____ Lower Upper Class

_____ Upper Class

In questions 23 - 40 place an ´X´ on a point on the scale that best indicates or describes your feelings. The ´X´ may be placed anywhere along the scale.

23. How do you feel about the job or position you currently have?

|--------------|--------------|--------------|-------------|

LIKE VERY MUCH	LIKE	NEUTRAL	DISLIKE	DISLIKE VERY MUCH

24. How much satisfaction do you derive from being a member of the Armed Services?

|--------------|--------------|--------------|-------------|

VERY SATISFIED	SATISFIED	BORDERLINE	UNSATISFIED	VERY UNSATISFIED

25. Computers are necessary in today´s society.

|--------------|--------------|--------------|-------------|

DECIDEDLY AGREE	SLIGHTLY AGREE	NO OPINION DON´T KNOW	SLIGHTLY DISAGREE	DECIDEDLY DISAGREE

26. How would voice recognition make a computer operator´s job?

|--------------|--------------|--------------|-------------|

MUCH EASIER	SOMEWHAT EASIER	NO OPINION	MORE DIFFICULT	MUCH MORE DIFFICULT

27. How would voice recognition equipment affect information processing or data input?

SAVE A LOT OF TIME	SAVE SOME TIME	NO OPINION DON´T KNOW	TAKES MORE TIME	TAKES A LOT MORE TIME

28. If voice recognition can save time, it would allow a keyboard operator to do other jobs.

DECIDEDLY AGREE	SLIGHTLY AGREE	NO OPINION DON´T KNOW	SLIGHTLY DISAGREE	DECIDEDLY DISAGREE

29. Describe the use of voice recognition equipment.

VERY EASY TO USE	EASY TO USE	NO OPINION	DIFFICULT TO USE	VERY DIFFICULT TO USE

30. What do you think of voice recognition equipment for use in Military Command Centers?

VERY PRACTICAL	SOMEWHAT PRACTICAL	NO OPINION DON´T KNOW	SOMEWHAT IMPRACTICAL	VERY IMPRACTICAL

31. How much previous computer experience have you had?

ALOT OF EXPERIENCE	CONSIDERABLE EXPERIENCE	SOME EXPERIENCE	VERY LITTLE EXPERIENCE	NO EXPERIENCE

32. What is your previous experience with voice recognition equipment?

`|--------------|--------------|--------------|-------------|`

VERY MUCH	MUCH	SOME	A LITTLE	NONE

33. How would additional experience with voice recognition equipment affect recognition accuracy?

`|--------------|--------------|--------------|-------------|`

MUCH IMPROVEMENT	SOME IMPROVEMENT	NO OPINION	A LITTLE IMPROVEMENT	NO IMPROVEMENT

34. How do you feel when a misrecognition occurs?

`|--------------|--------------|--------------|-------------|`

STRONGLY LIKE	LIKE	NEUTRAL	DISLIKE	STRONGLY DISLIKE

35. How do you feel when a non-recognition ('beep') occurs?

`|--------------|--------------|--------------|-------------|`

STRONGLY LIKE	LIKE	NEUTRAL	DISLIKE	STRONGLY DISLIKE

36. How do you feel when a recognition occurs?

`|--------------|--------------|--------------|-------------|`

STRONGLY LIKE	LIKE	NEUTRAL	DISLIKE	STRONGLY DISLIKE

37. Describe your participation in this experiment.

|--------------|--------------|--------------|--------------|

EXTREMELY COOPERATIVE	MODERATELY COOPERATIVE	COOPERATIVE	SOMEWHAT UNCOOPERATIVE	VERY UNCOOP-ERATIVE

38. How would you describe your participating in this type of experimentation?

|--------------|--------------|--------------|--------------|

STRONGLY LIKE	LIKE	NEUTRAL	DISLIKE	STRONGLY DISLIKE

39. What is your current physical condition?

|--------------|--------------|--------------|--------------|

OUTSTANDING	GOOD	AVERAGE	FAIR	POOR

40. If voice recognition does save time and allows YOU to be assigned other tasks, how often would YOU want to use it?

|--------------|--------------|--------------|--------------|

ALWAYS	FREQUENTLY	NOW AND THEN	SELDOM	NEVER

APPENDIX B

USER QUESTIONNAIRE #2

NAME:________________________________ SUBJECT#:____________

INSTRUCTIONS:

The purpose of this questionaire is to obtain information from you regarding physical characteristics, personal background, and opinions pertaining to voice recognition equipment and its use. Your answers will assist in determining whether personal and/or physiological traits contribute to effective utilization of voice recognition equipment.

The questions include multiple choice, YES/NO, rating scale and short answer (one or two words ONLY!) types. Appropriate guidance accompanies each question or block of questions.

Your name is NOT required but is requested in order to ease the necessary correlation of your replies with your results in the experimentation. If you desire anonymity, please respond with your subject number only. Please respond truthfully. Check your questionaire after completion to insure you've completed all the questions.

Thank-you for your assistance in this experiment.

In questions 1 - 3, provide either a one or two word response, or place an 'X' by the appropriate answer.

1. Have you ever had one or more of the following speech impediments and/or impairments?

_____ Articulation (difficulty in pronouncing vowels and/or consonants)

_____ Voice (irregularities in the larynx)

_____ Cleft lip and/or lip palate

_____ Cerebral palsy

_____ Stuttering

_____ Hearing impairments

_____ Aphasia

_____ Congenital speech defects (due to birth/pregnancy)

_____ Retardation

_____ None of the above

2. Have you ever received speech therapy from either a subsidized (free) clinic, private speech therapist, or through the public school system?

_____ YES

_____ NO

3. Have you ever received voice training or taken singing lessons?

_____ YES (How many years? _______)

_____ NO

In questions 4 - 15 place an 'X' on a point on the scale that best indicates or describes your feelings. The 'X' may be placed anywhere along the scale.

4. How would voice recognition make a computer operator's job?

|--------------|--------------|--------------|--------------|

MUCH EASIER	SOMEWHAT EASIER	NO OPINION	MORE DIFFICULT	MUCH MORE DIFFICULT

5. How would voice recognition equipment affect information processing or data input?

|--------------|--------------|--------------|--------------|

SAVE A LOT OF TIME	SAVE SOME TIME	NO OPINION DON'T KNOW	TAKES MORE TIME	TAKES A LOT MORE TIME

6. If voice recognition can save time, it would allow a keyboard operator to do other jobs.

|--------------|--------------|--------------|--------------|

DECIDEDLY AGREE	SLIGHTLY AGREE	NO OPINION DON'T KNOW	SLIGHTLY DISAGREE	DECIDEDLY DISAGREE

7. Describe the use of voice recognition equipment.

|--------------|--------------|--------------|--------------|

VERY EASY TO USE	EASY TO USE	NO OPINION	DIFFICULT TO USE	VERY DIFFICULT TO USE

8. If voice recognition does save time and allows YOU to be assigned other tasks, how often would YOU want to use it?

```
 |--------------|--------------|--------------|--------------|

ALWAYS      FREQUENTLY    NOW AND THEN      SELDOM        NEVER
```

9. How would additional experience with voice recognition equipment affect recognition accuracy?

```
 |--------------|--------------|--------------|--------------|

MUCH           SOME         NO OPINION     A LITTLE          NO
IMPROVEMENT  IMPROVEMENT                 IMPROVEMENT  IMPROVEMENT
```

10. How do you feel when a misrecognition occurs?

```
 |--------------|--------------|--------------|--------------|

STRONGLY       LIKE          NEUTRAL         DISLIKE     STRONGLY
LIKE                                                     DISLIKE
```

11. How do you feel when a non-recognition ('beep') occurs?

```
 |--------------|--------------|--------------|--------------|

STRONGLY       LIKE          NEUTRAL         DISLIKE     STRONGLY
LIKE                                                     DISLIKE
```

12. How do you feel when a recognition occurs?

¦--------------¦--------------¦--------------¦--------------¦

STRONGLY LIKE	LIKE	NEUTRAL	DISLIKE	STRONGLY DISLIKE

13. Describe your participation in this experiment.

¦--------------¦--------------¦--------------¦--------------¦

EXTREMELY COOPERATIVE	MODERATELY COOPERATIVE	COOPERATIVE	SOMEWHAT UNCOOPERATIVE	VERY UNCOOPERATIVE

14. How would you describe your participating in this type of experimentation?

¦--------------¦--------------¦--------------¦--------------¦

STRONGLY LIKE	LIKE	NEUTRAL	DISLIKE	STRONGLY DISLIKE

15. What do you think of voice recognition equipment for use in Military Command Centers?

¦--------------¦--------------¦--------------¦--------------¦

VERY PRACTICAL	SOMEWHAT PRACTICAL	NO OPINION DON'T KNOW	SOMEWHAT IMPRACTICAL	VERY IMPRACTICAL

APPENDIX C

SELF-EVALUATION QUESTIONNAIRE

NAME ___________________________ DATE_________ SUBJECT# ____

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement that indicates how you GENERALLY feel. There are no right or wrong answers. Please do not spend too much time on any one statement, but give the answer which seems to describe how you GENERALLY feel.

1 = ALMOST NEVER
2 = SOMETIMES
3 = OFTEN
4 = ALMOST ALWAYS

1.	I feel pleasant	1	2	3	4
2.	I tire quickly	1	2	3	4
3.	I feel like crying	1	2	3	4
4.	I wish I could be as happy as others seem to be	1	2	3	4
5.	I am losing out on things because I can´t make up my mind soon enough	1	2	3	4

6.	I feel rested	1	2	3	4
7.	I am "calm, cool, and collected"	1	2	3	4
8.	I feel that difficulties are piling up so that I cannot overcome them	1	2	3	4
9.	I worry too much over something that really doesn´t matter	1	2	3	4
10.	I am happy	1	2	3	4
11.	I am inclined to take things hard	1	2	3	4
12.	I lack self confidence	1	2	3	4
13.	I feel secure	1	2	3	4
14.	I try to avoid facing a crisis or difficulty	1	2	3	4
15.	I feel blue	1	2	3	4
16.	I am content	1	2	3	4
17.	Some unimportant thought runs through my mind and bothers me	1	2	3	4
18.	I take disappointments so keenly that I can´t put them out of my mind	1	2	3	4
19.	I am a steady person	1	2	3	4
20.	I get in a state of tension or turmoil as I think over my recent concerns and interests	1	2	3	4

SCORING KEY
for the
A-TRAIT EVALUATION

1.	4	3	2	1
2.	1	2	3	4
3.	1	2	3	4
4.	1	2	3	4
5.	1	2	3	4
6.	4	3	2	1
7.	4	3	2	1
8.	1	2	3	4
9.	1	2	3	4
10.	4	3	2	1
11.	1	2	3	4
12.	1	2	3	4
13.	4	3	2	1
14.	1	2	3	4
15.	1	2	3	4
16.	4	3	2	1
17.	1	2	3	4
18.	1	2	3	4
19.	4	3	2	1
20.	1	2	3	4

APPENDIX D

SELF-EVALUATION QUESTIONNAIRE

NAME ____________________________ DATE_________ SUBJECT# ____

DIRECTIONS: A number of statements which people have used to describe themselves are given below. Read each statement and then circle the appropriate number to the right of the statement that indicates how you feel RIGHT NOW -- AT THIS VERY MOMENT. There are no right or wrong answers. Please do not spend too much time on any one statement, but give the answer that best describes your PRESENT feelings.

1 = NOT AT ALL

2 = SOMEWHAT

3 = MODERATELY SO

4 = VERY MUCH SO

1.	I feel calm	1	2	3	4
2.	I feel secure	1	2	3	4
3.	I am tense	1	2	3	4
4.	I am regretful	1	2	3	4
5.	I feel at ease	1	2	3	4
6.	I feel upset	1	2	3	4
7.	I am presently worrying over possible misfortunes	1	2	3	4

8.	I feel rested	1	2	3	4
9.	I feel anxious	1	2	3	4
10.	I feel comfortable	1	2	3	4
11.	I feel self-confident	1	2	3	4
12.	I feel nervous	1	2	3	4
13.	I am jittery	1	2	3	4
14.	I feel "high strung"	1	2	3	4
15.	I am relaxed	1	2	3	4
16.	I feel content	1	2	3	4
17.	I am worried	1	2	3	4
18.	I feel over-excited and "rattled"	1	2	3	4
19.	I feel joyful	1	2	3	4
20.	I feel pleasant	1	2	3	4

SCORING KEY
for the
A-STATE EVALUATION

1.	4	3	2	1
2.	4	3	2	1
3.	1	2	3	4
4.	1	2	3	4
5.	4	3	2	1
6.	1	2	3	4
7.	1	2	3	4
8.	4	3	2	1
9.	1	2	3	4
10.	4	3	2	1
11.	4	3	2	1
12.	1	2	3	4
13.	1	2	3	4
14.	1	2	3	4
15.	4	3	2	1
16.	4	3	2	1
17.	1	2	3	4
18.	1	2	3	4
19.	4	3	2	1
20.	4	3	2	1

APPENDIX E

UTTERANCE LIST: TRAINING WEEK - WEEK#1

WORD#	UTTERANCE	CRT PROMPT
000	THREE	THREE
001	EUROPE	EUROPE
002	MOVE IT LEFT	MOVE IT LEFT
003	CARRIAGE RETURN	CARR RETURN
004	LOGOUT	LOGOUT
005	COMMAND	COMMAND
006	STRAIT OF HORMUZ	STR OF HMRZ
007	TIME	TIME
008	KOREA	KOREA
009	ZERO	ZERO
010	CHANGE DIRECTORY TO POOCK	C DIR TO PK
011	ALPHA	ALPHA
012	POSITIVE	POSITIVE
013	IDENTIFICATION	IDNTFICATION
014	LAUNCH	LAUNCH
015	RELOCATE	RELOCATE
016	DELTA	DELTA
017	TASK FORCE COMMANDER	TSK FRC CDR
018	KILO	KILO
019	LOGIN YELLEN	LOGIN YELLEN
020	ECHO	ECHO
021	NOVEMBER	NOVEMBER
022	TWO	TWO
023	UNITED STATES	UNITED STS
024	FOUR	FOUR
025	BRAVO	BRAVO
026	PLACE A CIRCLE ON MOSCOW	PL A CIR MOS
027	ENEMY DETECTION	EN DETECTION
028	PROCEED	PROCEED
029	ROMEO	ROMEO
030	FLIGHT CONTROLLER	FLT CTLR
031	SEVEN	SEVEN
032	GROUND CONTROL APPROACH	GND CTL APPR
033	REPORT	REPORT
034	AIRFIELD NAME	AFLD NAME
035	LIMA	LIMA
036	AVAILABLE	AVAILABLE
037	MESSAGE	MESSAGE
038	SATELLITE	SATELLITE
039	SHOOT	SHOOT
040	YANKEE	YANKEE
041	AFFIRMATIVE	AFFIRMATIVE

042	CHARLIE	CHARLIE
043	TORPEDO	TORPEDO
044	FIVE	FIVE
045	OPERATIONS PLAN	OPNS PLAN
046	OFFENSE	OFFENSE
047	UP IN DETAIL	UP IN DETAIL
048	NINE	NINE
049	PROBABILITY OF DETECTION	PROB OF DETN
050	NEUTRAL	NEUTRAL
051	JULIETT	JULIETT
052	SPEED	SPEED
053	UNIFORM	UNIFORM
054	SENSOR	SENSOR
055	TANGO	TANGO
056	CLOSE OUT CHARLIE	CLS OUT CHRL
057	LOAD THE GANN	LD THE GANN
058	OSCAR	OSCAR
059	NORTH ATLANTIC MAP	N ATL MAP
060	PACIFIC DATA BASE	PAC DAT BASE
061	HUMAN FACTORS	HUM FACTORS
062	FOXTROT	FOXTROT
063	SOVIET	SOVIET
064	DEFENSE	DEFENSE
065	ONE	ONE
066	INDIA	INDIA
067	ADVANTAGES	ADVANTAGES
068	GOLF	GOLF
069	CANCEL	CANCEL
070	ZULU	ZULU
071	NEGATIVE	NEGATIVE
072	PLOT ALL SUBMARINES	PLT ALL SUBS
073	XRAY	XRAY
074	REFUEL	REFUEL
075	AUTOMATIC RECOGNITION	AUTO RECOG
076	QUEBEC	QUEBEC
077	TRACK ENEMY	TRACK ENEMY
078	LEVEL TWO	LEVEL TWO
079	COURSE	COURSE
080	JOINT TASK FORCE	JT TSK FRC
081	SIX	SIX
082	WHISKEY	WHISKEY
083	ATTACK	ATTACK
084	SIERRA	SIERRA
085	MANEUVER DELAY	MNUVR DELAY
086	DISTANCE	DISTANCE
087	EXECUTE	EXECUTE
088	EIGHT	EIGHT
089	VICTOR	VICTOR
090	MEDITERRANEAN MAP	MED MAP
091	SEA OF JAPAN	SEA OF JAPN
092	POPPA	POPPA
093	FILE TRANSFER PROTOCOL	FL TNSFR PRO

094	ALTITUDE	ALTITUDE
095	HOTEL	HOTEL
096	NUKE THEM TILL THEY GLOW	NUKE EM
097	ACCAT TITLE	ACCAT TITLE
098	MIKE	MIKE
099	MISSILE	MISSILE

APPENDIX F

UTTERANCE LIST: WEEK #2

WORD#	UTTERANCE
000	MISSILE
001	MIKE
002	ACCAT TITLE
003	NUKE THEM TILL THEY GLOW
004	HOTEL
005	ALTITUDE
006	FILE TRANSFER PROTOCOL
007	POPPA
008	SEA OF JAPAN
009	MEDITERRANEAN MAP
010	VICTOR
011	EIGHT
012	EXECUTE
013	DISTANCE
014	MANEUVER DELAY
015	SIERRA
016	ATTACK
017	WHISKEY
018	SIX
019	JOINT TASK FORCE
020	COURSE
021	LEVEL TWO
022	TRACK ENEMY
023	QUEBEC
024	AUTOMATIC RECOGNITION
025	REFUEL
026	XRAY
027	PLOT ALL SUBMARINES
028	NEGATIVE
029	ZULU
030	CANCEL
031	GOLF
032	ADVANTAGES
033	INDIA
034	ONE
035	DEFENSE
036	SOVIET
037	FOXTROT
038	HUMAN FACTORS
039	PACIFIC DATA BASE
040	NORTH ATLANTIC MAP
041	OSCAR

042	LOAD THE GANN
043	CLOSE OUT CHARLIE
044	TANGO
045	SENSOR
046	UNIFORM
047	SPEED
048	JULIETT
049	NEUTRAL
050	PROBABILITY OF DETECTION
051	NINE
052	UP IN DETAIL
053	OFFENSE
054	OPERATIONS PLAN
055	FIVE
056	TORPEDO
057	CHARLIE
058	AFFIRMATIVE
059	YANKEE
060	SHOOT
061	SATELLITE
062	MESSAGE
063	AVAILABLE
064	LIMA
065	AIRFIELD NAME
066	REPORT
067	GROUND CONTROL APPROACH
068	SEVEN
069	FLIGHT CONTROLLER
070	ROMEO
071	PROCEED
072	ENEMY DETECTION
073	PLACE A CIRCLE ON MOSCOW
074	BRAVO
075	FOUR
076	UNITED STATES
077	TWO
078	NOVEMBER
079	ECHO
080	LOGIN YELLEN
081	KILO
082	TASK FORCE COMMANDER
083	DELTA
084	RELOCATE
085	LAUNCH
086	IDENTIFICATION
087	POSITIVE
088	ALPHA
089	CHANGE DIRECTORY TO POOCK
090	ZERO
091	KOREA
092	TIME
093	STRAIT OF HORMUZ

094	COMMAND
095	LOGOUT
096	CARRIAGE RETURN
097	MOVE IT LEFT
098	EUROPE
099	THREE

APPENDIX G

UTTERANCE LIST: WEEK #3

WORD#	UTTERANCE
000	CARRIAGE RETURN
001	STRAIT OF HORMUZ
002	ZERO
003	POSITIVE
004	RELOCATE
005	KILO
006	NOVEMBER
007	FOUR
008	ENEMY DETECTION
009	FLIGHT CONTROLLER
010	REPORT
011	AVAILABLE
012	SHOOT
013	CHARLIE
014	OPERATIONS PLAN
015	NINE
016	JULIETT
017	SENSOR
018	LOAD THE GANN
019	PACIFIC DATA BASE
020	SOVIET
021	INDIA
022	CANCEL
023	PLOT ALL SUBMARINES
024	AUTOMATIC RECOGNITION
025	LEVEL TWO
026	SIX
027	SIERRA
028	EXECUTE
029	MEDITERRANEAN MAP
030	FILE TRANSFER PROTOCOL
031	NUKE THEM TILL THEY GLOW
032	MISSILE
033	MOVE IT LEFT
034	COMMAND
035	KOREA
036	ALPHA
037	LAUNCH
038	TASK FORCE COMMANDER
039	ECHO
040	UNITED STATES
041	PLACE A CIRCLE ON MOSCOW

042	ROMEO
043	GROUND CONTROL APPROACH
044	LIMA
045	SATELLITE
046	AFFIRMATIVE
047	FIVE
048	UP IN DETAIL
049	NEUTRAL
050	UNIFORM
051	CLOSE OUT CHARLIE
052	NORTH ATLANTIC MAP
053	FOXTROT
054	ONE
055	GOLF
056	NEGATIVE
057	REFUEL
058	TRACK ENEMY
059	JOINT TASK FORCE
060	ATTACK
061	DISTANCE
062	VICTOR
063	POPPA
064	HOTEL
065	MIKE
066	EUROPE
067	LOGOUT
068	TIME
069	CHANGE DIRECTORY TO POOCK
070	IDENTIFICATION
071	DELTA
072	LOGIN YELLEN
073	THREE
074	TWO
075	BRAVO
076	PROCEED
077	SEVEN
078	AIRFIELD NAME
079	MESSAGE
080	YANKEE
081	TORPEDO
082	OFFENSE
083	PROBABILITY OF DETECTION
084	SPEED
085	TANGO
086	OSCAR
087	HUMAN FACTORS
088	DEFENSE
089	ADVANTAGES
090	ZULU
091	XRAY
092	QUEBEC
093	COURSE

094	WHISKEY
095	MANEUVER DELAY
096	EIGHT
097	SEA OF JAPAN
098	ALTITUDE
099	ACCAT TITLE

APPENDIX B

DATA COLLECTION FORM

NAME: ________________________ SEX: M F SUBJECT #: _______

RANK: ___________ DAY/TIME: ______________[TRIALS 1-2]

______________[TRIALS 3-4]

______________[TRIALS 5-6]

WEEK#: 1 2 3

MICROPHONE: _____EXPERIENCED _____NON-EXPERIENCED

TRAINING: _____SUPERVISED _____NON-SUPERVISED

UTTERANCE	TRIAL #					
	1	2	3	4	5	6
THREE						
EUROPE						
MOVE IT LEFT						
CARRIAGE RETURN						
LOGOUT						
COMMAND						
STRAIT OF HORMUZ						
TIME						
KOREA						
ZERO						
CHG DIR TO POCCK						

ALPHA						
POSITIVE						
IDENTIFICATION						
LAUNCH						
RELOCATE						
DELTA						
TASK FORCE CMDR						
KILO						
LOGIN YELLEN						
ECHO						
NOVEMBER						
TWO						
UNITED STATES						
FOUR						
BRAVO						
PL CIRCLE ON MOSCOW						
ENEMY DETECTION						
PROCEED						
ROMEO						
FLIGHT CONTROLLER						
SEVEN						
GRND CTRL APPROACH						
REPORT						
AIRFIELD NAME						
LIMA						

AVAILABLE						
MESSAGE						
SATELLITE						
SHOOT						
YANKEE						
AFFIRMATIVE						
CHARLIE						
TORPEDO						
FIVE						
OPERATIONS PLAN						
OFFENSE						
UP IN DETAIL						
NINE						
PROB OF DETECTION						
NEUTRAL						
JULIETT						
SPEED						
UNIFORM						
SENSOR						
TANGO						
CLOSE OUT CHARLIE						
LOAD THE GANN						
OSCAR						
NORTH ATLANTIC MAP						
PACIFIC DATA BASE						

HUMAN FACTORS							
FOXTROT							
SOVIET							
DEFENSE							
ONE							
INDIA							
ADVANTAGES							
GOLF							
CANCEL							
ZULU							
NEGATIVE							
PLOT ALL SUBMARINES							
XRAY							
REFUEL							
AUTO RECOGNITION							
QUEBEC							
TRACK ENEMY							
LEVEL TWO							
COURSE							
JOINT TASK FORCE							
SIX							
WHISKEY							
ATTACK							
SIERRA							
MANEUVER DELAY							

DISTANCE						
EXECUTE						
EIGHT						
VICTOR						
MEDITERRANEAN MAP						
SEA OF JAPAN						
POPPA						
FILE TNSFR PROTOCOL						
ALTITUDE						
HOTEL						
NUKE TILL THEY GLOW						
ACCAT TITLE						
MIKE						
MISSILE						

DATA REDUCTION

# NON-RECOGNITIONS						
# MIS-RECOGNITIONS						
# TOTAL ERRORS						

APPENDIX I

MASTER LIST OF UTTERANCES

1. ONE SYLLABLE UTTERANCES (15)

ONE
TWO
THREE
FOUR
FIVE
SIX
EIGHT
NINE
GOLF
MIKE
LAUNCH
TIME
SHOOT
SPEED
COURSE

2. TWO SYLLABLE UTTERANCES (35)

EUROPE
LOGOUT
ZERO
SEVEN
ALPHA
BRAVO
CHARLIE
DELTA
ECHO
FOXTROT
HOTEL
KILO
LIMA
OSCAR
POPPA
QUEBEC
TANGO
VICTOR
WHISKEY
XRAY
YANKEE

ZULU
COMMAND
REPORT
OFFENSE
DEFENSE
ATTACK
PROCEED
CANCEL
MESSAGE
DISTANCE
NEUTRAL
MISSILE
SENSOR
REFUEL

3. THREE SYLLABLE UTTERANCES (20)

MOVE IT LEFT
SOVIET
JOINT TASK FORCE
NOVEMBER
JULIETT
ROMEO
SIERRA
INDIA
UNIFORM
KOREA
NEGATIVE
POSITIVE
EXECUTE
AIRFIELD NAME
ALTITUDE
RELOCATE
LOAD THE GANN
LEVEL TWO
SATELLITE
TORPEDO

4. FOUR SYLLABLE UTTERANCES (14)

CARRIAGE RETURN
LOGIN YELLEN
STRAIT OF HORMUZ
UNITED STATES
FLIGHT CONTROLLER
AVAILABLE
AFFIRMATIVE
UP IN DETAIL

CLOSE OUT CHARLE
HUMAN FACTORS
ADVANTAGES
TRACK ENEMY
SEA OF JAPAN
ACCAT TITLE

5. UTTERANCES GREATER THAN OR EQUAL TO 5 SYLLABLES (16)

MANEUVER DELAY
CHANGE DIRECTORY TO POOCK
IDENTIFICATION
TASK FORCE COMMANDER
PLACE A CIRCLE ON MOSCOW
GROUND CONTROL APPROACH
ENEMY DETECTION
NORTH ATLANTIC MAP
MEDITERRANEAN MAP
PROBABILITY OF DETECTION
OPERATIONS PLAN
PACIFIC DATA BASE
PLOT ALL SUBMARINES
AUTOMATIC RECOGNITION
FILE TRANSFER PROTOCOL
NUKE THEM TILL THEY GLOW

APPENDIX J

INDIVIDUAL SUBJECT RECOGNITION RATES

The following are mean error rates for each subject participating in the experiment. The data is partitioned to mirror the groups established in the overall experimental design and are expressed in percent error.

GROUP I	GROUP II
4.89	13.11
7.17	9.22
7.39	8.69
4.39	8.39
9.22	5.22
6.44	6.89
6.23	6.72
8.06	5.33
1.61	4.06
2.89	2.22
2.61	1.67

GROUP III	GROUP IV
4.06	10.11
2.11	15.17
.50	4.89
8.94	15.72
9.28	8.06
4.33	9.06
5.72	8.44
8.22	6.28
4.50	2.39
2.94	7.11
3.61	4.33

LIST OF REFERENCES

1. Kryter, K.D., "Speech Communication", in Kinkade, R.G. and VanCott, H.P., *Human Engineering Guide to Equipment Design*, pp. 162-223, McGraw-Hill, 1972.

2. Doddington, G.R. and Schalk, T.B., "Speech Recognition: Turning Theory to Practice," *IEEE Spectrum*, pp. 26-32, September 1981.

3. Klatt, D.H., "Review of the ARPA Speech Understanding Project," in Dixon, N.R. and Martin, T.B., *Automatic Speech and Speaker Recognition*, IEEE Press, 1979.

4. Lea, W. A., "Speech Recognition: Past, Present and Future," in Lea, W.A., *Trends in Speech Recognition*, pp. 55-59, Prentice-Hall Inc., 1980.

5. Lea, W.A., *Computer Recognition of Speech*, Speech Science Publications, 1982

6. Martin, T.B., "Practical Applications of Voice Input to Machines," in Dixon, N. R. and Martin, T. B., *Automatic Speech and Speaker Recognition*, IEEE Press, 1979.

7. White, George M., "Speech Recognition: A Tutorial Overview", *Computer*, pp. 40-53, May 1976.

8. Naval Postgraduate School Report NPS55-80-016, *Experiments With Voice Input for Command and Control: Using Voice Input to Operate a Distributed Computer Network*, by G.K. Poock, April 1980.

9. Naval Postgraduate School Report NPS55-82-028, *Use of Voice Recognition Equipment With Stenographer Masks*, by G.K. Poock, N.D. Schwalm, and E.F. Roland, October 1982.

10. Naval Postgraduate School Report NPS55-82-032 *Trying for Speaker Independence in the Use of Speaker Dependent Voice Recognition Equipment*, by G.K. Poock, N.D. Schwalm, B.J. Martin, and E.F. Roland, December 1982.

11. Chapanis, A. and Ochsman, R.B., "The Effects of 10 Communication Modes on the Behaviour of Teams During Co-Operative Problem-Solving," *International Journal Man-Machine Studies*, Vol.6, September 1974

12. Operating Instructions for Wireless Input, Threshold Technology Inc., 1977.

13. Naval Postgraduate School Report NPS55-81-016, Effect of Operator Mental Loading on Voice Recognition System Performance, by J.W. Armstrong and G.K. Poock, August 1981.

14. Rothberg, Michael, "Applying Automatic Speech Recognition to Data Entry", Mini-Micro Systems, pp. 153-162, November 1980.

15. "Spoken Words Drive a Computer," Business Week, pp. 368-361, 2 December 1972.

16. Beek, B., Hodge, D.C., and Neuberg, E.P., "An Assessment of the Technology of Automatic Speech Recognition for Military Applications," in Dixon, N.R. and Martin, T.B., Automatic Speech and Speaker Recognition, IEEE Press, 1979.

17. Jay, G.T., An Experiment in Voice Data Entry for Imagery Interpretation Reporting, Masters Thesis, Naval Postgraduate School, Monterey, California, March 1981.

18. Malarkey, T.R., An Investigation of the Applications of Voice Input/Output Technology in the COINS Network Control Center, Masters Thesis, Naval Postgraduate School, Monterey, California, 1982.

19 Hutchingson, R.D., New Horizons for Human Factors in Design, McGraw-Hill, 1981.

20. Robinson, A.L., "More People are Talking to Computers as Speech Recognition Enters the Real World", Science, v. 203, pp. 634-638, 16 February 1979.

21. Lea, W.A., What Causes Speech Recognizers to Make Mistakes, paper presented at Short Course for Computer Recognition of Speech, Sunnyvale, California, 6-7 December 1982.

22. Batchellor. M.P., Investigation of Parameters Affecting Voice Recognition Systems in C3 Systems, Masters Thesis, Naval Postgraduate School, Monterey, California, 1981.

23. Spielberger, C.D., ANXIETY Current Trends in Theory and Research, pp. 23-46, v. 1, Academic Press Inc, 1972.

24. Naval Postgraduate School Report NPS55-81-013, A Longitudinal Study of Computer Voice Recognition Performance and Vocabulary Size, by G.K. Poock, June 1981.

25. Threshold 600 User's Manual, Threshold Technology Inc., 1978.

26. US Army Research Institute for the Behavioral and Social Sciences, Questionnaire Construction Manual, 1976.

27. Spielberger, C.D., Gorsuch, R.L., and Lushene, R.E., STAI Manual, Consulting Psychologists Press, Inc., 1969.

28. Ryan, T. A., Joiner, B. L., and Ryan, B. F., MINITAB Student Handbook, Duxbury Press, 1979.

29. Bruning, J. L. and Kintz, B. L., Computational Handbook of Statistics, Scott, Foresman and Company, 1968.

30. Conover, W. J., Practical Nonparametric Statistics, Wiley, 1980.

31. Ott, L., An Introduction to Statistical Methods and Data Analysis, Duxbury Press, 1977.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Superintendent ATTN: Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Superintendent ATTN: Professor G. Poock, Code 55Pk Naval Postgraduate School Monterey, California 93940	86
4.	Superintendent ATTN: CDR C. Hutchins USN, Code 55Hu Naval Postgraduate School Monterey, California 93940	1
5.	Superintendent ATTN: Professor D. Neil, Code 55Ni Naval Postgraduate School Monterey, California 93940	1
6.	Superintendent ATTN: Code 012A Naval Postgraduate School Monterey, California 93940	1
7.	Superintendent ATTN: Code 39 Naval Postgraduate School Monterey, California 93940	1
8.	IBM Communications Products Division ATTN: David W. Davenport PO Box 12195 Dept D48/B632 Research Triangle Park, NC 27709	1

9. SINGER 1
Link Simulation Systems Division
ATTN: Dr. E. Scott Baudhuin
11800 Tech Road
Silver Springs, Maryland 20904

10. Naval Electronics Systems Center 1
ATTN: Frank Deckelman, Code 613
2511 Jefferson Davis Highway
Arlington, Virginia 20363

11. Chairman, C3 Academic Group 1
ATTN: Professor Michael Sovereign, Code 39
Naval Postgraduate School
Monterey, Ca 93940

12. Director 1
National Security Agency
ATTN: Ms Joanne B. Kim, Code H44
Fort George Meade, Maryland 20755

13. US Army War College 1
Department of War Gaming
ATTN: CPT(P) H. W. Yellen
Carlisle Barracks, Pennsylvania 17013

200782

Thesis
Y345 Yellen
c.1 A preliminary analysis of human factors affecting the recognition accuracy of a discrete word recognizer for C3 systems.

20 FEB 8[illegible] 53431

20[illegible]32

Thesis
Y345 Yellen
c.1 A preliminary analysis of human factors affecting the recognition accuracy of a discrete word recognizer for C3 system.

A preliminary analysis of human factors
3 2768 000 98825 7
DUDLEY KNOX LIBRARY